Mandated data sharing in hybrid marketplaces

Federico Navarra* Flavio Pino† Luca Sandrini‡

February 2023

Abstract

Considering a monopolistic hybrid platform, we investigate the effect of a mandated data sharing policy on market outcomes across different data functionalities (price discrimination and cost reduction) and different market structures (perfect and imperfect competition). We find that mandated data sharing has no effects on welfare if data can be used to price discriminate consumers who buy homogeneous goods, while mandated sharing of cost-reducing data improves welfare by lowering the average price in the markets. When goods are horizontally differentiated, data sharing with price discrimination (cost reduction) may instead negatively (positively) affect consumers. We argue that in markets where competition is softer, mandatory data sharing may damage the very agents it is intended to protect, namely consumers and (efficient) sellers.

Keywords: *hybrid platforms, data sharing, vertical integration.*
JEL codes: *D42, L12, L41.*

*Department of Economics and Management, University of Padova (Italy); federico.navarra.1@phd.unipd.it
†Polytechnic of Turin (Italy); flavio.pino@polito.it
‡Quantitative Social and Management Sciences, Faculty of Economics and Social Sciences, Budapest University of Technology and Economics, Budapest (Hungary); sandrini.luca@gtk.bme.hu

1 Introduction

The relevance of data in modern economies has constantly been increasing during the past years, primarily due to the importance of digital markets. Every interaction on digital platforms and websites is tracked and registered by companies, and large amounts of data are traded every moment. Recent estimates from the European Commission show that:

> "The value of the data economy for the EU27 has been estimated to have reached almost €400 billion in 2019 and €440 billion in 2021, with a year-on-year growth rate of 4.9% in 2021. The estimated share of overall impacts on GDP in the EU27 ranges from 3.1% in 2019 to 3.6% in 2021" (DATA Market Study 2021–2023, pg. 116)."[1]

Data is a core input factor for production processes, logistics, targeted marketing, smart products, and services. Also, they are fundamental to training Artificial Intelligence and refining algorithms. On top of that, data drive interoperability in interconnected environments and are expected to impact specific sectors such as mobility and healthcare drastically.

Data owners have a large competitive advantage over their market rivals. Hence, data are very relevant to competition and privacy authorities. Digital platforms may have the incentive to adopt potential anti-competitive practices, such as self-preferencing (Padilla et al., 2022) and bias-recommendation (Bourreau and Gaudin, 2022), or, more generally, they may abuse their dominant position.

To stay competitive, firms competing against digital platforms increasingly depend on timely access to relevant data and their ability to use those to develop new, innovative applications, services, and products. For these reasons, a widespread debate has emerged on whether – and under which conditions and legal bases – public intervention is required to ensure adequate and timely access to data. One of the proposed remedies is to mandate platforms to share with sellers and rival companies all or part of the consumers' data they possess. Data sharing is one of the pillars of the European strategy for data.[2] It is at the core of the Digital Market Act (DMA hereafter), the recently introduced EU competition law regulating large digital platforms' (gatekeepers) business conduct.[3]

This paper analyses the strategic interactions between a monopolistic platform and the many sellers operating within the digital marketplace. Moreover, we investigate the effects of mandated data sharing on market outcomes and social welfare when data can be used for different purposes (price discrimination or cost reduction). We focus on a setting where a digital platform mediates between many sellers and

[1] Available at `https://digital-strategy.ec.europa.eu/en/library/results-new-european-data-market-study-2021-2023`.

[2] Available at `https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020DC0066&from=EN`.

[3] In particular, article 6 of the DMA states that: *"The gatekeeper shall provide business users and third parties authorized by a business user, at their request, free of charge, with effective, high-quality, continuous and real-time access to, and use of, aggregated and non-aggregated data, including personal data, that is provided for or generated in the context of the use of the relevant core platform services or services provided together with, or in support of, the relevant core platform services by those business users and the end users engaging with the products or services provided by those business users".* (DMA, Art. 6.10), available at `https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32022R1925&from=EN`.

Table 1: Effect of data sharing on main market outcomes.

	Homogeneous Goods		Differentiated Goods	
	Price Disc.	Cost Red.	Price Disc.	Cost Red.
CS	= or ↓	↑	↓	↑
TW	=	↑	↑	↑

consumers. Sellers must pay a per-transaction fee to the platform to be included in the marketplace, whereas consumers do not pay any admission fee. The platform decides, in order, i) the size of the fee, ii) whether or not to directly produce some (or all) final goods and compete against the sellers, and iii), in case of entry, the price of each good it produces. This setup is consistent, as an example, with Amazon's product groups' referral fees. Amazon subdivides its marketplace into broad product groups such as "baby products" or "clothing and accessories". While these categories contain many sub-markets, Amazon sets a single per-transaction fee for every product group. Moreover, we assume the platform has an ex-ante data advantage against the sellers, meaning that it can use the data it owns exclusively. Mandated data sharing alters the interaction between the platform and sellers depending on the competition structure (homogeneous or differentiated goods).

Our main result is summarized in Table 1. Mandated data sharing has no effects on welfare if data can be used to price discriminate consumers and competition is in prices of homogeneous goods. Indeed, we show that the platform sets a per-transaction fee that makes all sellers less efficient than the platform itself. Hence, with or without data, the sellers can only set a price equal to their marginal costs and cannot exploit consumers' data. The platform may also decide to operate as anintermediary in some markets, letting some sellers without competitive pressure and free to perfectly price-discriminate consumers.

Instead, mandated sharing of cost-reducing data improves welfare by lowering the average price in the market. Sellers are unaffected, as price competition drives profits to zero, but the platform is better off as more markets are covered and revenues from intermediation increase. When goods are horizontally differentiated, data sharing may negatively affect consumers (price discrimination) and some sellers (cost reduction). In the former case, access to data enables sellers to price more efficiently and extract more surplus from the consumers, and the platform sets the per-transaction fee to temper competition with the sellers. In the latter case, efficient sellers have little gains from the cost-reducing data but suffer from the generalized increase in transaction fees. However, the aggregate effect is positive.

Our second result pertains to the pricing strategies adopted by the platforms and the sellers. We identify conditions under which the platform has the incentives to either *strategically lose* price competition (homogeneous goods) or to give up competing in some market segments (data sharing with differentiated goods). In the former case, the platform's strategy is to set a price the rival can undercut. In particular, the platform sets its price just above the rival's marginal costs. It does so because winning the price competition yields lower revenues than collecting the transaction fees. Similarly, in the latter case, the platform decides not to send any offer to some consumers as it would be less profitable than collecting the fee from the sellers. This result stems from the vertically integrated (i.e., hybrid) nature of the platform, which operates as an intermediary and a rival in the market. Earning revenues in both cases, the platform exploits its advantage and sets a per-transaction fee to lower the efficiency of rival sellers. Then, depending on whether the competitive price guarantees higher or lower net revenues than the transaction fee, the platform decides whether to win or to lose the competition.

Data is a complex concept that subsumes a multiplicity of information, uses, and functions. Indeed, when discussing data access, one has to consider the nature of data. A list of consumers may be used to send out discounts, tailored ads, and prices. Information about the production process of a specific good may be used to improve the efficiency of firms in producing it. Searches and reviews by consumers may be used to assess the size of the demand for a specific good and potential fallacies that bother consumers, limiting retail risks.

Hence, from an anti-trust perspective, the data's importance strongly depends on analysing the specificities of a given market and available data. Investigating an overall impact on market outcomes without explicitly accounting for the data type may limit the breadth of policy implications and provide little guidance to policy makers.

In recent years, many authors have investigated the effects of data on various aspects, such as market structure, competition, welfare, and privacy. However, most of these works have focused on a specific data effect in particular settings. This, in turn, has created a conundrum. While the impact of data is widely analyzed, the specificity of most studies makes it difficult to compare models and abstract more general insights.

In this paper, we compare different usages of data in different market structures, and we provide some guidance from a regulatory perspective about what one should expect from mandating data sharing. As mentioned earlier, we argue that in markets where competition is softer, mandatory data sharing may damage the very agents it is intended to protect, namely consumers and (efficient) sellers.

The paper is organized as follows: in the next section, we provide an overview of the existing literature on the topic and illustrate our main contributions. Section 3 presents the model with homogeneous goods. We analyze the effects of mandated data sharing when data allow firms to either price discriminate consumers 3.1 or to reduce their costs 3.2. In Section 4, we introduce horizontal differentiation: we repeat the analysis of the effects of mandated data sharing in sections 4.1 and 4.2. Finally, in Section 5 we compare our results with those of the related literature and draw concluding remarks.

2 Related literature

This paper contributes to four main strands of literature. The first one focuses on the effect of consumer data in digital economics. The use of data is widespread across every sector, thanks to their versatility;

typical uses include improving products or services quality and efficiency, personalisation, matching, and discriminating between different consumers groups or individuals. Recent surveys (Goldfarb and Tucker, 2019; Bergemann and Bonatti, 2019; Pino, 2022) have thus focused on categorizing both the types and uses of data, trying to extract broader insights that hold across different models. Two typical data functions are those of allowing price discrimination on consumers and of increasing the vertical differentiation between firms (either allowing for an improvement of the products or a reduction in their marginal cost of production).

Price discrimination has been observed in various markets: a typical example involves the use of geolocalization to tailor prices to different consumers (Mikians et al. (2012); Aparicio et al. (2021)). The literature has mostly focused on competition between informed firms, stemming from Thisse and Vives (1988) seminal work, and on the vertical relations between firms and a data seller (Montes et al. (2019); Bounie et al. (2021); Delbono et al. (2021); Abrardi et al. (2022)). The common insight of these models is that allowing all firms to obtain data benefits consumers while it harms both the firms and the data seller. Our main contribution highlights how, when vertical integration is introduced, mandating data sharing can benefit all actors, as firms can retain more profits due to their increase in competitiveness while the platform benefits from the overall increase in market efficiency.

Data can also help firms increase their products' value or decrease the associated costs. An example of the former is using consumer data to optimize product offerings (Campbell et al. (2015)) and versioning (Bhargava and Choudhary (2008)), while the latter can be due, among others, to a reduction in search, replication, and transportation costs (Goldfarb and Tucker (2019)). All these effects point towards an increase in the vertical differentiation between firms if one has more data than the other. Our analysis shows that mandating data sharing reduces vertical differentiation, increasing competition and benefiting consumers. Sellers are better off as the gap between them and the platform is reduced, and platform profits increases as it can charge a higher fee.

The second strand of literature focuses instead on information sharing. Information sharing has been extensively studied in the literature: Raith (1996) describes a general model that summarizes many existing models, to show the determinants of when and how firms are incentivized to disclose private information. The recent literature on digital economics is also gaining interest with regard to information sharing, with particular attention on consumer data. Krämer and Schnurr (2022) focus on market contestability with regard to data-rich incumbents, and explore the possible effects of policy interventions such as data siloing, data sharing and data portability. Regarding to e-commerce, they stress the importance of sellers' data portability, as this policy would allow sellers to grow without having to lock-in on a specific platform. Prüfer and Schottmüller (2021) study competition in data-driven market where data reduce the cost of quality production. Their model shows how mandated data sharing does not reduce the dominant firm's incentive to innovate and also eliminates the risk of market tipping. Krämer and Shekhar (2022) expand on this topic by analyzing how the aforementioned policy interventions impact competition between platforms, modeling the effect of data as an improvement in the user experience on the platforms and allowing platforms to compete as well as set their investment levels. In particular, they show that mandated data sharing can reduce innovation investment by platforms, which in turn can hurt consumers when data externalities are large. De Corniere and Taylor (2020) analyze the effects of data sharing by using a competition in utilities approach, finding sufficient conditions under which data sharing would be unambiguously pro-competitive. Liu et al. (2021) focus instead on a retail platform that hosts sellers and can strategically disclose information to them: they find that the platform has the incentive to disclose information only to a subgroup of sellers. While our work focuses on an exogenous shock that mandates complete information sharing instead of allowing for a strategic decision by the platform, as far as we know we are the first to allow the platform to vertically integrate, entering the downstream market and competing with sellers.

Related to the vertical integration aspect of our model, the third strand of literature concerns the classical questions regarding access pricing and sabotage. Indeed, our model resembles the typical setup of an upstream monopolist that controls an infrastructure and can choose to integrate downstream. Economides (1998) shows how an integrated monopolist has the incentive to degrade the quality of the downstream input, as to raise the costs of its rivals until they are driven out of the market. Beard et al. (2001) expand on this topic by showing that the upstream monopolist is always willing to expand downstream, but has the incentive to sabotage only when input price regulation is introduced. Our model presents a similar result: the platform can strategically use the per-transaction fee to increase downstream costs, allowing it to better compete against sellers. Moreover, in the case of price discrimination with heterogeneous goods, the platform sets the fee such that sellers opt to price as monopolists, increasing surplus extraction from consumers and, in turn, the platform's profits.

Finally, the third strand of literature concerns hybrid marketplaces – platforms that allow transactions between sellers and buyers and where the platform can become a seller's competitor. This literature is becoming pivotal in policy discussion, as tech giants such as Amazon and Apple are themselves hybrid marketplaces. Empirical evidence suggests that the downstream entry of the platform, sometimes referred to as *dual mode*, usually takes place in successful markets and leads sellers to reduce their growth efforts in the platform (Zhu and Liu (2018)). Moreover, sellers tend to increase the prices in the markets where the platform enters, while shifting their innovation investments elsewhere (Wen and Zhu (2019)). In particular, evidence suggests that complementary goods become the focus of innovation, as the platform entry usually expands the demand for that good (Foerderer et al. (2018)). From a theoretical perspective, the effects of a platform operating in dual mode are ambiguous. On the one hand, platform downstream entry could reduce sellers' market power and increase competition, benefiting consumers (Dryden et al. (2020); Etro (2021b)). Platform entry could also induce it to reduce its commission fees to further expand the market's reach (Etro (2021a)). On the other hand, a higher quality (or lower cost) of the platform's goods can incentivize it to increase its commission fees, ultimately harming consumers (Anderson and Bedre-Defolie (2021)). While our model assumes the presence of a hybrid marketplace, focusing instead on mandated data sharing, our results show that an increase in vertical differentiation by the platform leads indeed to an increase in the commission fee, which in turn can crowd out sellers.

3 Homogeneous goods

We consider a monopolistic platform which operates in dual mode, namely it intermediates between sellers and consumers while selling directly to consumers as well. We consider a mass of independent sellers ranked by their marginal costs of production c_s which we assume uniformly distributed between 0 and ∞. Each technology is allocated to only one seller, and each seller is a monopolist in his market. However, the platform can enter in a market where a seller is present. There is a unitary mass of consumers for every product space (namely for every seller) and consumers are uniformly distributed according to their willingness to pay $v \in [\underline{v}, \bar{v}]$ with $\underline{v} < \bar{v}$. Every consumer buys at most one unit of any given good.

Both the platform and sellers know the distribution of the willingness to pay but sellers cannot always price discriminate consumers. The platform intermediates between sellers and consumers by imposing a per-transaction fee f. Note how the platform sets a unique fee for all markets: our modeling choice is consistent with Amazon's practice of setting a specific fee per product group.[4] Our model thus aptly represents one of Amazon's product groups, which is composed of multiple sub-markets. Our analysis could be easily replicated in any one of Amazon product groups to find the optimal fee for each of them. Moreover, allowing the platform to operate fee discrimination within a product group would give it even more market power: our analysis thus describes a lower bound with respect to the welfare effects of the hybrid platform.

When the platform and a seller operate in the same product space, they compete in prices *à la Bertrand*. We assume that the platform is always able to serve customers with its products, thus it incurs a marginal production cost $c_p \in [0, \underline{v})$ which is equal across product spaces.[5]

Platform cost c_p is known by the platform and revealed to sellers when the platform enters their markets. In order to capture the fact that the platform exploits its data advantage for avoiding retail risk, we assume that the platform can produce good i only if seller i enters the market (joining the online marketplace).[6] In other words, we never consider the platform to be a monopolist on its marketplace for a given product space.

In line with the measures recently proposed by the Digital Markets Act (DMA), we focus our analysis on mandated data sharing, namely we do not let the platform to decide over its data sharing policy but it rather suffers a third-party decision on whether to share platform data with third-party sellers or not. We then compare the resulting outcomes with and without data sharing. As we observe in reality, data can serve several purposes and can improve seller efficiency in multiple ways, in our analysis we

[4]This type of fee is labeled by Amazon as a "referral fee": *Amazon charges a referral fee on each item sold. The amount varies depending on the product group.* See `https://sell.amazon.com/pricing#referral-fees`.

[5]The assumption of a single technology that the platform can use to produce every good allows us to generate asymmetries between the platform and the sellers. We are interested in analyzing a situation in which the platform could be more or less efficient than the relevant seller, depending on the sub-market they are operating in. One may argue that, for the model to be realistic, the platform should face different technologies in each market: this assumption, though, allows us to keep the model as simple and clear as possible.

[6]See https://www.reuters.com/investigates/special-report/amazon-india-rigging/ and Madsen and Vellodi (2022)

first consider data as a price discriminating tool and later as a cost reducing one.

The timing is the following:

1. The platform sets the per-transaction fee.

2. The platform selects which of the downstream markets to enter.

3. Price competition takes place.

The solution concept is the Sub-game Perfect Nash Equilibrium (SPNE).

3.1 Data allow price discrimination

We begin our analysis by considering data as an essential tool for price discrimination. In other words, if sellers can access data they are able to distinguish consumers and their actual willingness to pay, thus setting different prices. Otherwise, sellers set a unique price for all consumers. In our model, the platform always price discriminates since it enjoys all the data gathered on the marketplace. When the platform is forced to share its data, sellers price discriminate as well.

3.1.1 Platform data advantage (no data sharing)

Stage 1: Sellers are initially independent monopolists. After having observed the fee f set by the platform, they set prices in order to maximize their profits. Without data sharing, sellers know the distribution of consumers' willingness to pay but cannot tell what is the willingness to pay of each consumer, hence sellers set a unique price. Seller i sets p_{s_i} in order to maximize $\pi_{s_i} = (p_{s_i} - c_{s_i} - f)d(p_{s_i})$ where $d(p_{s_i}) = 1 - \frac{p_{s_i} - v}{\bar{v} - \underline{v}}$. The optimal seller i's price without data sharing is $p^*_{s_i} = \frac{\bar{v} + c_{s_i} + f}{2}$ such that the seller serves a share $d(p^*_{s_i}) = \frac{\bar{v} - c_{s_i} - f}{2(\bar{v} - \underline{v})}$ of market i.

Stage 2: The platform decides for each product market i whether to operate as intermediary (*pure agency*) or to enter the product market, thus acting also as a first-party seller (*dual mode*). In the former case, the platform earns a fee f on every transaction which takes place in market i, thus earning a profit of $\Pi_i = fd(p^*_{s_i})$ in each market in which it operates in pure agency. In the latter one instead, the platform enters the market and sets the price of the final good strategically: we show that the platform has two pricing strategies: i) to compete à la Bertrand against the seller and win the market ii) to strategically lose the price competition by setting a price just above the rival's marginal costs. In what follows, we will refer to the latter strategy as *constrained agency*, meaning that the platform still earns only the per-transaction fee f (platform profit from direct sales is still zero) but on a larger customer base. In fact, its price constrains the third-party seller's price at the competitive level.

Instead, under standard Bertrand competition the platform enjoys two revenue sources: the profit from intermediation and the one from direct sales. Which of the option prevails once the platform enters the market is the outcome that maximizes platform profit, which is formally defined as:

$$\Pi_i^D M = fd(p_{s_i}^{DM}) + (p_{p_i} - c_p)d(p_{p_i})$$

Stage 3: When the platform operates in *constrained agency*, it enters the market and loses the competition strategically. However, by doing so, it exerts competitive pressure on the seller's price. If the seller is more cost-efficient than the platform ($c_{s_i} < c_p$) it sets a price equal to $p_p = c_p$. Otherwise, the platform sets a price just above the marginal costs of the seller, so that the seller can always win the competition by setting a price equal to its own marginal costs $p_{s_i} = c_{s_i} + f$.

In the former case, given that $c_p < \underline{v}$ by assumption, the seller serves the entire market i and the platform earns $\Pi_i = f$, while in the latter case the seller serves $d(f + c_{s_i}) \leq 1$ and the platform earns $\Pi_i = fd(f + c_{s_i})$. Straightforward calculations prove that platform's profit under *constrained agency* is larger than the one under *pure agency* if $fd(p_{s_i}^*) < fd(f + c_{s_i})$ which is equivalent to:

$$f\left(\frac{\bar{v} - c_{s_i} - f}{2(\bar{v} - \underline{v})}\right) < f\left(\frac{\bar{v} - c_{s_i} - f}{(\bar{v} - \underline{v})}\right)$$

which holds as long as $c_{s_i} < \bar{v} - f$.

Notice that those markets in which this condition is satisfied are all the ones in which sellers are profitable. We define those markets as *active*. Indeed, if the maximum willingness to pay $\bar{v}$ was lower than the sellers' marginal cost, they would not make positive profits even as monopolists and they would not join the marketplace in the first place. It follows that in every market the platform prefers to enter rather than to operate as a pure intermediary.

Lemma 1. *In every active market, operating in pure agency is a dominated strategy for the platform, as entry allows it to earn strictly higher profits.*

The platform can win the Bertrand competition against the seller i only when it enjoys a lower marginal cost. Keeping in mind that i) the marginal costs of the sellers are endogenously determined by the fee set by the platform, but ii) the fee has already been set at the pricing stage, an efficient platform ($c_p < c_{s_i} + f$) can always set a price p_p equal to the seller's marginal costs and serve all the consumers with a willingness to pay $v \in [f + c_{s_i}, \bar{v} - f]$.[7] On top of this, using its data advantage, the platform can also fully extracts the surplus from those consumers with a low willingness to pay $v \in [\underline{v}, f + c_{s_i})$ and could not be served by the seller. Wrapping everything up, the profit of the platform in market i under Bertrand competition is:

$$\Pi_i^{DM} = (f + c_{s_i} - c_p)d(f + c_{s_i}) + \int_{\underline{v}}^{f+c_{s_i}} v\,dv \tag{1}$$

[7]Note that $\bar{v} - f$ is the upper-bound for sellers' marginal cost given that those with more expensive technologies would never be profitable and do not join the marketplace in the first place.

where the first element in equation (1) is the profit made by competing for the consumers that have a willingness to pay $v \geq c_{s_i} + f$, while the the second element represents the surplus extracted by the platform through price discrimination from the consumers that the seller in unable to serve.

Alternatively, the platform could potentially operate under *constrained agency*, setting its price just above the seller's marginal cost. As a result, the platform would not earn from selling downstream, but only via collecting per-transaction fees f. Moreover, given its data advantage, the platform could still extract surplus through price discrimination from the consumers that the seller is unable to serve.

One can see that Bertrand competition is superior to *constrained agency* when

$$(f + c_{s_i} - c_p)d(f + c_{s_i}) + \int_{\underline{v}}^{f+c_{s_i}} v\, dv \geq f d(f + c_{s_i}) + \int_{\underline{v}}^{f+c_{s_i}} v\, dv \tag{2}$$

which holds true for $c_{s_i} \geq c_p$.

Lemma 2. *In every active market i in which $c_{s_i} \geq c_p$, the platform compete à la Bertrand. Otherwise, the platform strategically loses the price competition and the seller sets $p_{s_i} = c_{s_i} + f$, and serves the entire market i.*

Platform total profit is defined as:

$$\Pi = \int_0^{c_p} f\, dc_{s_i} + \int_{c_p}^{\bar{v}-f} \left((f + c_{s_i} - c_p)d(f + c_{s_i}) + \int_{\underline{v}}^{f+c_{s_i}} v\, dv \right) dc_{s_i}, \tag{3}$$

where the first term represents the platform's profit from *constrained agency* while the second one captures the platform's profit from Bertrand competition. When sellers' marginal cost exceeds $\bar{v} - f$ they exit the marketplace since they cannot earn positive margins even with a monopoly price. Platform's total profit is concave in f since a fee increase would reduce the extensive margins (demand contraction) and increase the intensive ones (per-transaction revenues expansion), hence an interior solution for the optimal fee may exist.

Proposition 1. *The optimal fee is $f^* = c_p$, such that the platform is weakly more cost-efficient than every third-party seller.*

Proof. The result stems from standard profit-maximization of equation (3). □

Proposition 1 implies that each active seller makes zero profit, while consumer surplus is equal to:

$$CS = \int_0^{\bar{v}-c_p} \int_{c_p+c_{s_i}}^{\bar{v}} v - c_{s_i} - c_p\, dv\, dc_{s_i}$$

One can see that, with $f = f^*$, the marginal costs of the seller in any given market is $c_{s_i} + c_p$, which is always larger than or equal to c_p. This implies:

Proposition 2. *When data allow price discrimination, the equilibrium prices under no data sharing are the following. If $c_{s_i} \geq c_p$, the platform sets $p_p = c_p + c_{s_i} - \varepsilon$ and serves the entire market. If $c_{s_i} < c_p$, the platform sets $p_p = c_p + c_{s_i} + \varepsilon$ and strategically loses the price competition. The seller always sets $p_{s_i} = c_p + c_{s_i}$.*

Proof. Proposition 2 derives from using the optimal fee f^* in the results stated in Lemma 2. □

Propositions 1 and 2 describe the equilibrium of the game. In fact, i) if the platform is ex-ante more efficient than the seller ($c_p < c_{s_i}$), then $f = f^*$ implies that the platform always wins the price competition by charging $p_p = c_{si} + f^* = c_{si} + c_p$, which allows it to gain a net margin of c_s on every transaction. ii) If the seller is ex-ante more efficient than the platform ($c_p > c_{s_i}$), the latter enters the market and chooses to strategically lose the Bertrand competition, setting a price just above the seller's marginal costs $p_p = c_{s_i} + c_p + \varepsilon$, with ε arbitrarily small. Hence, the platform loses the competition and earns only from intimidation. Because it faces no costs of production, the available margin is $f^* = c_p > c_{s_i}$. In both cases, the final price is $c_{s_i} + c_p$, meaning that there is no difference between the two strategies on the extensive margin.

3.1.2 Data sharing

When the platform is forced to share the gathered data, third party sellers can perfectly price discriminate consumers, namely they know the exact willingness to pay of each consumer in their market. Nevertheless, the presence of the platform in every market constrains sellers' pricing strategies such that prices are equal to marginal costs with and without data sharing.

Indeed, if sellers were monopolists, thanks to price discrimination they would serve a share $d(f + c_{s_i})$ in market i and the platform would make a *pure agency* profit of $\Pi_i = fd(f + c_{s_i})$.[8] If the platform operates in *constrained agency* instead, it would make at least the same profit; the platform decides to strategically lose and sets the price of its product equal to $p_p = f + c_{s_i} + \varepsilon$. In this way it forces the seller to set $p_{s_i} = f + c_{s_i}$ such that it earns a profit $\Pi_i = fd(f + c_{s_i})$, which is equal to the *pure agency* one.

The platform then does not have strong incentives to adopt *pure agency* and may decide to always enter its own marketplace as first-party seller. In this case, it is easy to see that pricing strategies are the same with or without data sharing. As a result, the number of active firms in the marketplace does not change after data sharing and sellers make zero profit despite their ability to price discriminate consumers. Hence, consumer surplus does not change after the policy.

Lemma 3. *When the platform decides to enter as first-party seller in every market after data sharing, equilibrium pricing strategies are identical to the case without data sharing.*

As mentioned before, the platform may also opt to act as intermediary in those markets in which sellers are more cost-efficient. In this case, although the platform makes the same profit of the case

[8] A monopolist seller with data sharing serves every consumer with a willingness to pay higher than its marginal cost.

without data sharing and with data sharing when it always acts as first-party seller, efficient sellers ($c_{s_i} < c_p$) do not face any competitive pressure in their product market and are free to price discriminate consumers. It follows that: efficient sellers are better off, inefficient sellers are indifferent and consumers are worse off after policy implementation.

Lemma 4. *When the platform decides to act as intermediary in markets where it does not enjoy a cost advantage, sellers are able to price discriminate consumers.*

Given that the Platform is indifferent between *pure agency* and *constrained agency*, the condition that determines in which markets the platform operates in *dual mode* stays the same also after data sharing, namely $c_{s_i} > c_p$. The same applies to the cost threshold that determines seller entry in the marketplace which stays equal to $c_{s_i} = \bar{v} - f$.

Proposition 3. *When data allows price discrimination, if the platform decides to operate as intermediary in markets where $c_{s_i} < c_p$, data sharing makes: efficient sellers better off, inefficient sellers indifferent and consumers worse off, without affecting total welfare. Otherwise, data sharing is not effective since equilibrium prices and the equilibrium fee are unaltered.*

Proof. Proposition 3 derives from the discussion above. □

Given this result, one can see that a mandatory data sharing policy is unlikely to improve either consumer surplus or total welfare. The only change we observe is the increase in the profit of cost-efficient sellers to the detriment of consumers.
When data allows price discrimination, data sharing does not expand market entry as it does not make sellers more efficient. Therefore, in order to increase total welfare, competition should increase within existing markets but this is not possible given that platform and sellers already compete *à la Bertrand* without data sharing. In other words, data sharing cannot improve total welfare once perfect competition is in place. If any, data sharing tends to reduce the competitive pressure on efficient sellers, being highly detrimental for consumers.

3.2 Data allow cost reduction

We now turn to the case when data are used for reducing production costs. To keep the analysis simple, we assume that the platform enjoys a data advantage such that it has a marginal cost $c_p^{CR} = c_p(1 - r)$, where $r \in [0, 1]$ is exogenous. The data-driven cost advantage can be related to versioning (Bhargava and Choudhary, 2008), or to a reduction in replication and transportation costs (Goldfarb and Tucker, 2019).[9] When the platform has to share its data, every seller i can lower its marginal cost by lowering their marginal cost from c_{s_i} to $c_{s_i}^{CR} = c_{s_i}(1 - r)$.

[9] To keep the model as simple as possible, we are thus assuming that the platform and the sellers are equally effective in using data to decrease their costs.

3.2.1 Platform data advantage (no data sharing)

Stage 1 and **stage 2** of the game stay the same as in the previous price discrimination case and Lemma 1 holds also with data for cost reduction, while outcomes and their conditions change in **stage 3**. The platform can undercut seller i only when it enjoys a lower marginal cost. It follows that in market i the platform prefers competing à la Bertrand than strategically losing the price competition when the following condition holds

$$(f + c_{s_i} - c_p^{CR})d(f + c_{s_i}) \geq fd(f + c_{s_i}). \tag{4}$$

Expression (4) holds when $c_{s_i} > c_p^{CR}$. Viceversa, keeping in mind that $c_p^{CR} \leq c_p < \underline{v}$, the platform operates in *constrained agency* and the seller serves the entire market.

Lemma 5. *In every active market i in which $c_{s_i} \geq c_p^{CR}$, the platform competes à la Bertrand. Otherwise, the platform strategically loses the price competition and the seller sets $p_{s_i} = c_{s_i} + f$, and serves the entire market i.*

Platform profit is defined as

$$\Pi = \int_0^{c_p^{CR}} f\, dc_{s_i} + \int_{c_p^{CR}}^{\bar{v}-f} (f + c_{s_i} - c_p^{CR})d(f + c_{s_i})\, dc_{s_i} \tag{5}$$

Following the same maximization strategy as for price discrimination, we derive the optimal fee:

Proposition 4. *The optimal fee is $f^\dagger = c_p^{CR}$, such that the platform is weakly more cost-efficient than every third-party seller.*

Proof. The result stems from standard profit-maximization of equation (5). □

Using the optimal fee in the aggregate sellers' profits, one can see that

$$\Sigma\pi = \int_0^{c_p^{CR}} (c_{s_i} + c_p^{CR} - c_{s_i} - c_p^{CR})\, dc_{s_i} = 0 \tag{6}$$

and consumer surplus is

$$CS = \int_0^{\bar{v}-c_p^{CR}} \int_{c_p^{CR}+c_{s_i}}^{\bar{v}} (v - c_{s_i} - c_p^{CR})\, dv\, dc_{s_i} \tag{7}$$

Proposition 5. *When data allow cost reduction, the equilibrium prices under no data sharing are the following. If $c_{s_i} \geq c_p^{CR}$, the platform sets $p_p = c_p + c_{s_i} - \varepsilon$ and serves the entire market. If $c_{s_i} < c_p^{CR}$, the platform sets $p_p = c_p + c_{s_i} + \varepsilon$ and strategically loses the price competition. The seller always sets $p_{s_i} = c_p^{CR} + c_{s_i}$.*

Proof. Proposition 5 derives from using the optimal fee $f^\dagger$ in the results stated in Lemma 5. □

3.2.2 Data sharing

When data sharing is mandated, sellers reduce their marginal costs to $c_{s_i}^{CR} = c_{s_i}(1-r)$.

By the same logic as under no data sharing, it is trivial to observe that the platform would never operate in *pure agency.* Hence, upon entry, the platform has to decide whether to compete à la Bertrand or to adopt *constrained agency* strategy. The former is superior when

$$(f + c_{s_i}^{CR} - c_p^{CR})d(f + c_{s_i}^{CR}) \geq fd(f + c_{s_i}^{CR}), \tag{8}$$

which holds when $c_{s_i} > \frac{c_p^{CR}}{1-r} = c_p$.

Lemma 6. *In every active market i, the platform enters the market and operates in dual mode. Moreover, given the fee f, when $c_{s_i} \geq c_p$, the platform compete à la Bertrand. Otherwise, the platform strategically loses the price competition and the seller sets $p_{s_i} = c_{s_i}^{CR} + f$ and serves the entire market i.*

Lemma 6 simply states that, when data allow cost reduction, mandatory data sharing makes the platform decide to strategically lose (win) price competition in more (less) markets than in the case without data sharing.

The platform profit becomes

$$\Pi = \int_0^{c_p} f \, dc_{s_i} + \int_{c_p}^{\frac{\bar{v}-f^*}{1-r}} (f + c_{s_i}^{CR} - c_p^{CR})d(f + c_{s_i}^{CR}) \, dc_{s_i} \tag{9}$$

Proposition 6. *The optimal fee is $f^\dagger = c_p^{CR}$, such that the platform is weakly more cost-efficient than every third-party seller.*

Proof. The result stems from standard profit-maximization of equation (9). □

Using f^* in the sellers' payoff, one can see that the aggregate sellers' profits are

$$\Sigma\pi = \int_0^{c_p} (c_{s_i}^{CR} + c_p^{CR} - c_{s_i}^{CR} - c_p^{CR}) \, dc_{s_i} = 0 \tag{10}$$

while consumer surplus is

$$CS = \int_0^{\frac{\bar{v}-c_p^{CR}}{1-r}} \int_{c_p^{CR}+c_{s_i}}^{\bar{v}} (v - c_{s_i}^{CR} - c_p^{CR}) \, dv \, dc_{s_i} \tag{11}$$

It follows that:

Proposition 7. *When data allow cost reduction, the equilibrium prices under mandated data sharing are the following. If $c_{s_i} \geq c_p$, the platform sets $p_p = c_p^{CR} + c_{s_i}^{CR} - \varepsilon$ and serves the entire market. If $c_{s_i} < c_p$, the platform sets $p_p = c_p^{CR} + c_{s_i} + \varepsilon$ and strategically loses the price competition. The seller always sets $p_{s_i} = c_p^{CR} + c_{s_i}^{CR}$.*

Proof. Proposition 7 derives from using the optimal fee $f^\dagger$ in the results stated in Lemma 6. $\square$

Data sharing expands the cost threshold for seller entry from $\bar{v}-f^*$ to $\frac{\bar{v}-f^*}{1-r}$, which increases with the share of cost reduction r. New sellers that join the marketplace are less efficient than the platform and therefore make zero profits. Nevertheless, sellers' entry also allows consumers that would have stayed out of the market without data sharing to make transactions. Therefore, consumer surplus increases as a result of the increased number of active markets. On top of that, data sharing also implies a generalized reduction of prices. Indeed, one can see that, in the already existing markets i with $c_{s_i} \in [0, \bar{v} - c_p^{CR}]$, the price level decreases by $c_{s_i}r$. This price effect entails a second positive effect of data sharing on consumer surplus
By taking the difference between the consumer surplus with and without data sharing we get:

$$\Delta CS = \frac{(\bar{v} - \underline{v})(\bar{v} - c_p^{CR})(\underline{v} - c_p^{CR})r}{2(1-r)} > 0$$

Although the platform makes smaller profits in every market i with $c_{s_i} \in [0, \bar{v} - c_p^{CR}]$, the demand expansion entailed by data sharing more than compensate the loss in per- transaction margins in each market. Formally, the difference between the platform profits with and without data sharing writes as:

$$\Delta\Pi = \frac{r\bar{v}^2}{2(1-r)} - r^2(\bar{v} + c_p^{CR}) > 0 \quad \forall\ 0 < r \leq 1$$

Define the (utilitarian) social welfare as the sum of aggregate profits and consumer surplus. Because sellers are unaffected by mandated data sharing, while consumers and the platform are strictly better off, it is trivial to observe that the effect of data sharing on social welfare is positive.

Proposition 8. *Mandated data sharing is welfare-enhancing and never harms consumers. Moreover, sellers are unaffected by the policy while the platform strictly benefits from it.*

Proof. Results derived from the discussion above. $\square$

4 Differentiated goods

The analysis of competition with homogeneous goods captures some interesting features of the platform economy. Indeed, on platforms like Amazon Marketplace, in many product categories, sellers offer the same identical good to many consumers and compete for prominence (e.g., to appear in the Amazon *buy box*). To do so, they engage in Bertrand-like price competition, trying to increase their quality-price ratio (Ciotti and Madio, 2022).

However, digital marketplaces are also populated by sellers and buyers who offer and demand different varieties of given goods. When consumers' demand for variety is high, the substitutability between different producers decreases, and competition softens (Hagiu, 2009).

In this section, we replicate the analysis by adopting a different model specification. By doing so, we aim to further investigate the strategic interaction between the platform and sellers when consumers have heterogeneous valuations of the goods offered in the market.

We show that, in this framework, data-sharing may indeed have anti-competitive effects (it harms consumers) if data allow price discrimination, or it may harm some sellers if data allow cost reduction. In general, data sharing is welfare improving, but not in the sense of Pareto.

Moreover, and noteworthy, we emphasize that data sharing can be harmful to the economic agents it is originally intended to favor (consumers and sellers), whereas it always benefits the platform.

Model set up. Consider a digital marketplace owned by a platform. The marketplace groups together a unit mass of markets denoted by $\mathcal{I}$. In each market $i \in \mathcal{I}$, a seller (s^i) and the platform (p) compete in prices for horizontally differentiated goods. In what follows, we will sometimes refer to the seller and the platform together as to *firms*.

We assume that in each market there is a continuum of consumers uniformly distributed on the $[0-1]$ Hotelling line. They consume at most one unit of either the good sold by the seller or the one sold by the platform. A consumer located in $x \in [0,1]$ derives constant utility $u > 1$ from consuming either of the two goods and pays a price p_k, where $k = s, p$. Also, she suffers a mismatch disutility $t|z_k - x|$ from consuming a variety that is not her favorite one, where $t > 0$ is the transportation cost, and z_k is the location of the variety consumed ($z_s = 0$ and $z_p = 1$). Throughout the model, we assume that t is sufficiently large so that no firm can under any circumstance cover the entire market alone. In each market i, the consumers' utility functions are:

$$\begin{aligned} U_s^i &= u - p_s^i - t\,x \\ U_p^i &= u - p_p^i - t(1-x) \\ U_{no}^i &= 0 \end{aligned}$$

where the subscript $_{no}$ stands for *no consumption*.

In any market $i \in \mathcal{I}$, the seller and the platform compete in prices and sell two varieties of one good. We assume production does not involve any fixed cost, but the seller active in i produces at a marginal cost $c_s^i \in [0,1]$. Instead, the platform produces all final goods at the same constant marginal cost $c_p \in [0,1]$. Similarly to the model with homogeneous goods, we assume that sellers are heterogeneous in their cost of productions ($c_s^i \neq c_s^{-i}$). Moreover, c_s^i is uniformly distributed between 0 and 1, and each technology is allocated to only one seller. In other words, we can map each technology c_s^i to a market i.

In addition to the marginal costs of production, all sellers have to pay the same per-transaction fee $f > 0$ to the platform in order to sell on the marketplace. Consequently, the payoffs of the two firms are:

$$\pi_s^i = D_s^i(p_s^i - c_s^i - f), \qquad \pi_p^i = D_p^i(p_p^i - c_p) + D_s^i\, f$$

where D_k^i indicates the demand of each firm $k = s, p$ in market i and includes all consumers who derive larger utility from consuming the good produced by k than by $-k$ or than not consuming at all. Notice that the platform earns revenues from per-transaction fees paid by the seller.

In each market, the seller must earn net revenues to stay active. It must set a price that is at least as high as the marginal costs of production, which are determined by the cost parameter c_s^i and, crucially, by the fee f set by the platform. By adjusting f, the platform can alter the price of the seller and the market demands, which are determined by the locations of the indifferent consumers.

Market configurations. Three possible market configurations may emerge. First, the locations of the consumers indifferent between buying from either the platform or the seller and not buying at all are such that $\tilde{x}_{s,no} > \tilde{x}_{p,no}$. In this case, keeping in mind that the location of the two firms are $z_s = 0$ and $z_p = 1$, there exists a consumer $\tilde{x}_{s,p} \in (\tilde{x}_{s,no}, \tilde{x}_{p,no})$ who is indifferent between buying from the seller or the platform and derives positive utility in both cases. This is the standard Hotelling duopoly (hd) case with full market coverage. Firms compete in the product market and prices are strategic complements.

Second, the locations of the indifferent consumers are such that $\tilde{x}_{s,no} < \tilde{x}_{p,no}$. In this case, consumers located in $x \in (\tilde{x}_{s,no}, \tilde{x}_{p,no})$ prefers not buying at all and the market is not covered. This is the local monopolies scenario (lm), in which firms do not compete against each other and prices are not set strategically.

Finally, the locations of the consumers indifferent between buying from firm i and not buying at all are such that $\tilde{x}_{s,no} = \tilde{x}_{p,no} = \tilde{x}_{s,p}$. In other words, there exists a range of values of v such that the platform and the seller can achieve higher profits by pricing like monopolists, while the market is fully covered. This scenario is referred in the literature as *monopolistic duopoly* (md) case (Thépot, 2007; Bacchiega et al., 2021). Prices are strategic substitutes and firms adjust them strategically to ensure the market is just covered. Differently from the Hotelling duopoly market configuration, in this one, the indifferent consumer derives zero utility from consuming either goods.

The timing of the game is the following: $t = 0$) the policy maker introduces a data-sharing policy. If there is no data-sharing, the platform uses data exclusively. Otherwise, all sellers can also use data. $t = 1$) the platform sets a single per-unit linear fee which is the same in all markets. All sellers have to pay it to be allowed to sell their goods on the marketplace. $t = 2$) Given the fee f, sellers and the platform set prices simultaneously in each market. If data allow the data owners to price discriminate consumers, as they know their exact locations on the Hotelling line, we model price competition as in Montes et al. (2019), and Bounie et al. (2021), among others.[10] $t = 3$) Consumers observe the prices

[10]When the platform holds information about consumers' location, but the sellers don't, a well-known problem is the existence of a pure strategy Bertrand-Nash equilibrium (see Rhodes and Zhou, 2022, p.25). In order to ensure equilibrium existence, we assume that personalized price schedules are set only after uniform prices are set. Consistently, Amazon allegedly shows higher prices to Amazon Prime subscribers but compensates them with discounted services such as free shipping. See `https://www.consumeraffairs.com/news/lawsuit-alleges-amazon-charges-prime-members-for-free-shipping-031414.html`

and decide if and what they consume.[11] The solution concept is Subgame Perfect Nash Equilibrium, and the game is solved by backward induction.

For sake of clarity, in what follows, we will omit the apex ^{i} when doing so does not create confusion.

4.1 Data allow price discrimination

Consider any given market i. Data ownership allows the platform to operate first-degree price discrimination to all consumers. The price set by the uninformed seller is uniform for all consumers, while the platform can offer tailored prices to each consumer.

We model tailored prices as the prices that make consumers indifferent between buying from the platform and the best available alternative option (i.e., buying from the seller at a uniform price or not buying at all):

$$p_p^{TC}(x) = p_s - t + 2tx, \qquad p_p^{TM}(x) = u + tx - t,$$

The apex ^{TC} stands for *Tailored under Competition*, suggesting a price that makes the consumer indifferent between buying from p or s. Instead, the apex ^{TM} indicates a price *Tailored under Monopoly* that makes the consumer indifferent between buying from the platform or not buying at all. We relegate all mathematical steps and proofs to the Appendix.

4.1.1 Platform data advantage (no data sharing)

First, suppose that data sharing is not mandated, and thus only the platform can operate first-degree price discrimination. The platform offers a price schedule $p_p(x)$. Hence, in any given market i its payoff function adjusts as follows:

$$\pi_p^i(f) = \begin{cases} \int_{\tilde{x}_{s,p}}^{\tilde{x}_{s,no}} \; p_p^{TC}(x) - c_p \, dx + \int_{\tilde{x}_{s,no}}^{1} \; p_p^{TM}(x) - c_p \, dx + D_s^i \, f & \text{if} \quad \tilde{x}_{s,p}^i \geq \tilde{x}_{p,no}^i \\ \int_{\tilde{x}_{p,no}}^{1} \; p_p^{TM}(x) - c_p \, dx + D_s^i \, f & \text{if} \quad \tilde{x}_{s,p}^i < \tilde{x}_{p,no}^i \end{cases} \tag{12}$$

Equation (12) shows that the platform surplus extraction crucially depends on whether the targeted consumer does or does not consider purchasing both goods. Indeed, a consumer located after $\tilde{x}_{s,no}^i$ would never purchase the good from the seller. Hence, the platform sets a tailored price that extracts all of the consumer's surplus. Instead, a consumer who is located before $\tilde{x}_{s,no}^i$, would compare the prices of the two firms and consume accordingly. Thus, the platform's ability to extract surplus is limited by the presence of the rival.

The platform operates in all markets. Hence, its aggregate profit function can be written as

$$\Pi_p = \int_0^1 \pi_p^i(f) \, dc_s^i$$

[11] Consistently with the literature, we assume consumers do not observe more than one price per firm. In other words, if a consumer observes personalized price by a firm, she cannot compare it with a uniform price by the same firm.

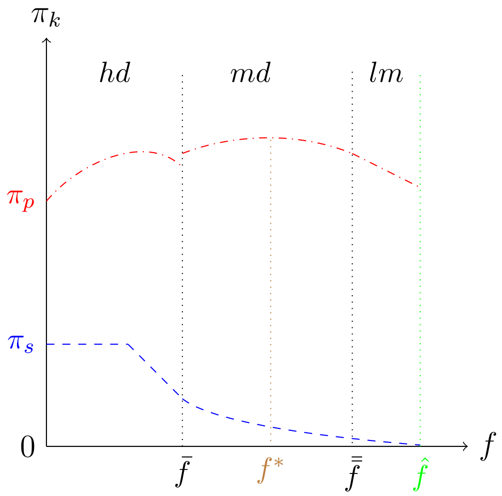

Figure 1: The equilibrium market configuration. $\forall\ c_s^i, c_p \in [0,1]$ the optimal platform sets a fee f^* such that firms act as monopolistic duopoly (md). The red dashdotted curve is the profit function of the platform, whereas the blue dashed curve is the profit function of the seller. Thresholds $\bar{f}$ and $\bar{\bar{f}}$ sort between Hotelling duopoly (hd) and monopolistic duopoly, and between monopolistic duopoly and local monopolies (lm), respectively. Above $\hat{f}$ the seller leaves the market.

As mentioned above, depending on the size of f, which is given at the pricing stage, the two firms can be competing in a Hotelling duopoly, a monopolistic duopoly, or they can operate as local monopolies. In the Appendix, we identify the cut-off values of f such that the different market configurations emerge, and solve the game under each configuration them to derive the equilibrium in every market. Moreover, we prove that:

Proposition 9. *Assume there is no data sharing. The platform set a fee*

$$f^* = \frac{2(u + 2(t + c_p))) - 3}{10}$$

such that the equilibrium market configuration is monopolistic duopoly (md) in every market i. The prices are:

$$p_s^i = \frac{u + c_s^i + f^*}{2}; \quad p_p^i(x) = \frac{u + c_s^i + f^* + 2t(2x - 1)}{2}$$

Proof. See the Appendix. □

The intuition for the result is the following: on the one hand, the platform wants to set a high fee to extract more profits from the seller, while on the other hand, a fee too high would result in some consumers not buying any good. By setting a fee such that all markets fall under the monopolistic duopoly case, the platform achieves three results: i) it extracts all surplus from its captive consumers, ii) it softens competition by charging monopoly pricing, and iii) it ensures that every market is fully covered.

Figure 1 illustrates the main result in Proposition 9. The cut-off value $\bar{f}$, which is depends on c_p and c_s^i, represents the level of fee such that, if $f = \bar{f}$, the seller is indifferent between reacting competitively to the price schedule of the platform or charging a monopoly price. Noteworthy, this is not the same threshold that makes the platform indifferent between competitive and monopoly pricing. Thus, in equilibrium, platform profits exhibit a discontinuity at $\bar{f}$.

4.1.2 Data sharing

Consider now data sharing is mandated so that sellers can also operate first-degree price discrimination. Depending on the market configuration, we can sort consumers into three groups. The first one includes those consumers that can only be profitably targeted by the seller. Consequently, the seller offers them its tailored price $p_s^{TM}(x)$ end extracts all the surplus from them. Similarly, the second group includes those consumers that can only be profitably targeted by the platform. To them, the platform offers a tailored price $p_p^{TMk}(x)$. Finally, the third group includes those consumers who can be reached by both the seller and the platform, and are thus contested. Since both firms can price discriminate, they can technically compete *à la Bertrand* for each consumer of the third group. To do so, they set a tailored price $p_k^{TC}(x)$.

The intense competition to *conquer* contested consumers generates ambiguous incentives on the platform. Because of the transportation costs, the firms have to offer consumers prices that are decreasing in the preference mismatch (i.e., the distance between consumers and firms' locations). However, the platform earns f from every consumer who purchases the seller's variety. Consequently, the platform may adopt sophisticated pricing strategies to *regulate* competition with the seller for consumers in the third group (i.e., contestable ones). More in detail, the platform may not be able to extract at least $c_p + f$ via personalized pricing from consumers whose preferred variety is sufficiently far from variety p. In this case, allowing the seller to serve those consumers may be the most profitable choice.

To understand this seemingly counterintuitive result, consider a platform with a marginal cost c_p. If there exist some consumers who are contested, then the platform's standard strategy is to offer $p_p^{TC}(x)$ and undercut the rival. However, if $p_p^{TC}(x) < c_p + f$, the net revenues the platform earns from winning the price competition is $p_p^{TC}(x) - c_p < f$. Instead, by giving up those consumers and allowing the seller to serve them, the platform earns f without producing anything (no costs involved). This strategy is similar, but not equal, to the *constrained agency* strategy highlighted in the model for homogeneous goods. In both cases, the platform prefers strategically losing price competition because it is more efficient to let the seller serve the market. The main difference, however, is that in this case, the platform does not *contrain* the seller's pricing strategy to the marginal cost of production. Instead, by giving up the competition, it allows the seller to fully extract the surplus from those consumers.

Obviously, giving up the price competition is a viable option if and only if those consumers are contested (third group). Otherwise, if they cannot be profitably targeted by the seller, the platform earns nothing from not offering them a tailored price. We define $\tilde{x}_{p2,no} > \tilde{x}_{p,no}$ the location of the last

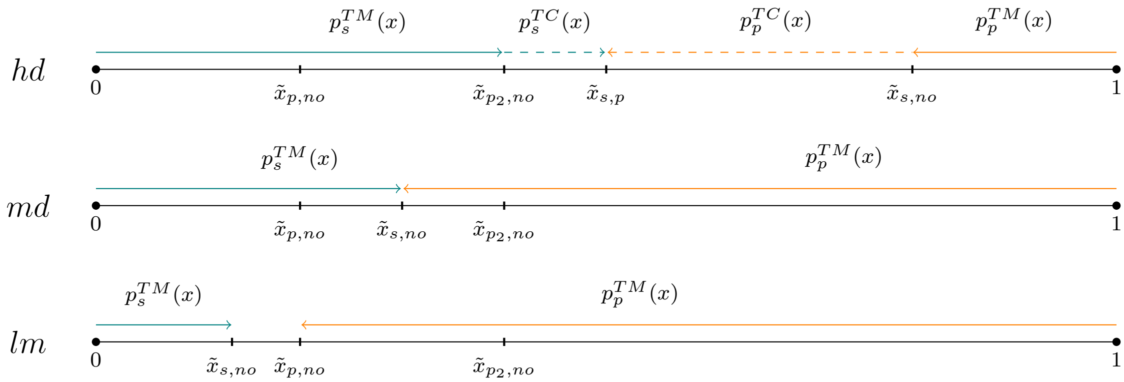

Figure 2: Market configuration and pricing strategies. When $f > \bar{\bar{f}}_{ds}$, firms are local monopolies (*lm*). If $f \in [\bar{f}_{ds}, \bar{\bar{f}}_{ds}]$, firms operate in monopolistic duopoly. Finally, if $f < \bar{f}_{ds}$, firms compete in a Hotelling duopoly. For illustrative reasons, the diagram shows the case $c_p = c_s^i$.

contested consumer that the platform can conquer with a price $c_p + f$. Figure 2 describes graphically the combination of market configuration and pricing strategies of the two firms.

In the Appendix, we identify the cut-off values of f such that the different market configurations emerge, and solve the game under each configuration to derive the equilibrium in every market. Moreover, we prove that:

Proposition 10. *Assume there is data sharing. The platform sets a fee*

$$f_{ds}^* = \frac{3u - t - c_p - 1}{3}$$

such that the equilibrium market configuration is monopolistic duopoly (md) in every market i. The equilibrium prices are:

$$p_s^{TM}(x) = u - t\,x \qquad p_p^{TM}(x) = u - t(1 - x);$$

Proof. See the Appendix. □

Corollary 1. *The platform strategically gives up competition for consumers located in $x^i \in [\tilde{x}^i_{p,no}, \tilde{x}^i_{s,no}]$. It does so because the transaction fee earned by allowing the seller to serve those consumers is larger than the surplus the platform could directly extract through personalized pricing.*

Proof. The proof of Corollary 1 stems from the following considerations. First, define $\tilde{x}_{p_2,no}$ as the last consumer from whom the platform can extract at least $c_p + f$. This consumer is necessarily closer to the platform than the consumer in $\tilde{x}_{p,no}$, by construction. It follows that, if the consumer indifferent between the seller's good and the zero payoff is located in $\tilde{x}_{s,no} \in [\tilde{x}_{p,no}; \tilde{x}_{p_2,no}]$, then the platform earns $f > p_p^{TC}(x \in [\tilde{x}_{p,no}, \tilde{x}_{s,no}]) - c_p$ by renouncing to compete. □

The intuition behind the results stated in Proposition 10 and Corollary 1 resembles the one described for the scenario with no data sharing: the platform can maximize its profits by ensuring full market coverage while simultaneously softening competition with the seller. As to the equilibrium fee, we find that f^*_{ds} can be lower or higher than f^*, depending on the parameters. In particular, when c_p is relatively low, the platform can serve most of the consumers in all markets. If this is the case and given the efficiency of the seller, the platform can strategically set a larger fee to raise its rival costs, without uncovering the market. Instead, if c_p is relatively high, the platform may not be able to cover a large section of the market. Lowering the fee enables the sellers to serve more consumers and allows the platform to keep earning revenues from all consumers in the market. One may also notice that, due to data sharing, the equilibrium fee decreases faster in c_p than without data sharing. Indeed, data allow the sellers to price more efficiently and to cover larger sections of the market.

Finally, from a welfare perspective, we find that:

Proposition 11. *Data sharing exerts a negative effect on consumers, who are strictly worse off, and a positive effect on firms' payoffs, which strictly increase. Total welfare increases as the latter effect dominates the former.*

Proof. See the Appendix. □

Proposition 11 states that all consumers are worse off under mandated data sharing, whereas firms are better off. On the one hand, the platform can extract more profits through the per-unit fee, while still using it to avoid direct competition with sellers. On the other hand, sellers become more efficient in extracting surplus and thus enjoy higher profits. Instead, all consumers obtain zero surplus in equilibrium. In fact, by controlling the fee, the platform sets up a monopolistic duopoly with the seller, and both of them can perfectly extract surplus through price discrimination, leaving consumers with no residual utility. This result is novel in the literature. Most models stemming from Thisse and Vives (1988) seminal work highlight how allowing both competitors to price discriminate leads to price wars, which largely benefit consumers. Interestingly, we show that, by strategically using the fee, the platform can avoid price wars altogether, leading to a fully covered market under monopolistic pricing.

4.2 Data allow cost reduction

Let us now turn to the case where data allow firms to produce the final good more efficiently. We impose the following modification to the model presented in the previous section: 1) The cost of the platform is given by $c_p^{CR} = c_p(1-r)$, where $r \in [0,1]$ indicates the intensity of the cost-reducing effect of data. 2) There is no price discrimination, meaning that both the sellers and the platform, in each market, set a single price for their goods. 3) Data sharing allows the seller to produce the final good at a reduced cost $c_s^{i,CR} = c_s^i(1-r)$ in all markets

4.2.1 Platform data advantage (no data sharing)

First, consider the scenario where the platform does not share its database with sellers. In all markets, it produces the final good at a marginal cost $c_p^{CR} \in [0,1]$. Hence, the platform has a data advantage,

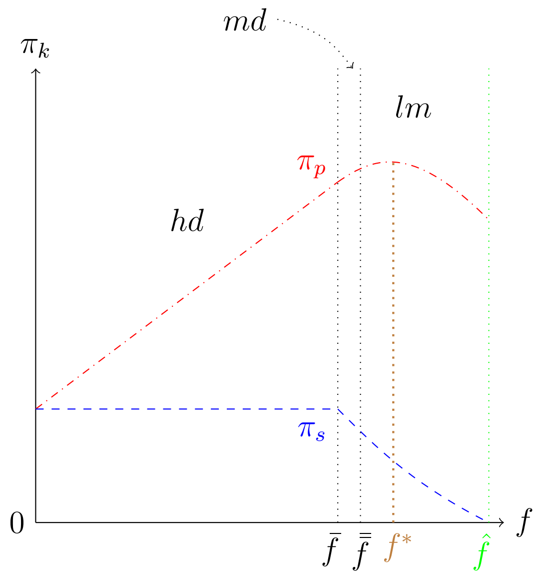

Figure 3: The equilibrium market configuration. $\forall\ c_s,\ c_p,\ r \in (0,1)$ the optimal platform sets a fee that is such that firms act as local monopoly (lm). The red dashdotted curve is the profit function of the platform, whereas the blue dashed curve is the profit function of the seller. Thresholds $\bar{f}$ and $\bar{\bar{f}}$ sort between Hotelling duopoly (hd) and monopolistic duopoly, and between monopolistic duopoly and local monopolies (lm), respectively. Above $\hat{f}$ the seller leaves the market. For illustrative reasons, the graph shows the case $c_p(1-r) = c_s$.

as it can exploit the cost-reducing potential of data exclusively. The two strategic variables that the platform chooses are the transaction fee (linear per unit) that applies to all sellers and is the same across all markets, as well as the market-specific price. Instead, sellers only decide the market-specific price of their final goods. We relegate all mathematical steps and proofs to the appendix. The model solves as a standard Hotelling game.

In the appendix we prove the following proposition:

Proposition 12. *Assume there is no data sharing. The platform set a fee*

$$f^{\dagger} = \frac{2u-1}{4}$$

such that the equilibrium market configuration is local monopolies (lm) in every market i. The equilibrium prices are:

$$p_s^i = \frac{u + c_s^i + f^{\dagger}}{2} \qquad p_p^i = \frac{u + c_p^{CR}}{2};$$

Proof. See the Appendix. □

Proposition 12 shows that an interior optimal f always exists and it is located in the region where the market is only partially covered and firms are local monopolies. This result is graphically illustrated in Figure 3. The intuition behind it is the following: if the market is fully covered (hd), the firms compete in prices and the seller transfers its entire marginal costs to the consumers via pricing. This includes also the per-unit transaction fee that sellers must pay to the platform. Consequently, the platform has the incentive to raise the fee, at least to the point when the market starts uncovering. This is the well-known *increase rival's costs* strategy. Because prices are strategic complements, the platform can increase the fee (hence earning larger revenues from intermediation) as well as its price (hence earning larger profits from direct sales).

However, the seller cannot keep transferring the fee to consumers indefinitely, as they have a limited willingness to pay for the two goods. In particular, median consumers who suffer the highest utility loss from preference mismatch would eventually decide not to purchase any good if the prices go above their consumption value, net of the transportation costs. We define $\bar{f}$ as the value of the fee above which the pricing strategy described here stops working.

Above $\bar{f}$, the platform needs to adjust its price in order to prevent the market from uncovering. As charging a larger fee implies that the seller increases its price, the platform has to lower react strategically by lowering its own price. By doing so, the platform ensures that the consumers left over by the seller due to high prices do not abandon the market. This pricing strategy can be sustained up to the point when the platform finds lowering the price less profitable than allowing some consumers to abandon the market, which occurs when $f > \bar{\bar{f}}$.

Finally, above $\bar{\bar{f}}$, the two firms are local monopolies. In this region, increasing the fee has two opposite effects on the revenues from intermediation: on the one hand, it raises the intensive margin (the platform earns more per transaction); on the other hand, increasing the fee decreases the extensive margin (less consumers purchase from the sellers in each market). There exists a level $f^{\dagger}$ above which the negative effect on the extensive margin dominates the positive effect on the intensive margin.

One may notice that this setting resembles the problem of horizontal licensing of a process innovation (or technology transfer) as illustrated by Kabiraj and Lee (2011), among others. However there is one main difference that is worth mentioning. In the literature on licensing, the size of the royalty (f in our setting) is set so that the market is fully covered. Indeed, because it is impossible to find a royalty rate that simultaneously i) satisfy the conditions of partial market coverage, ii) it is profitable for the licensor, and iii) it is profitable for the licensee, the equilibrium royalty must be the highest one such that, in equilibrium, all consumers purchase a unit of one good.

In our setting, however, the seller pays the fee to be active on the platform, not to get the cost-reducing data. In other words, it is as the platform was licensing a fundamental input, not a non-drastic innovation. Hence, the exit option for the seller is the zero payoff, which strongly relaxes condition iii) above. As a result, a fee that implies local monopolies can be an equilibrium and, in fact, it turns out to be the only equilibrium of the game.

4.2.2 Data sharing

Assume now a policy maker imposes data sharing so that the platform has to hand over its database to all sellers in each market. The marginal cost function of the seller in market i becomes $c_s^{i,CR}$, as it is now able to improve its efficiency using the cost-reducing effect of data. The payoffs of the two firms in each market (dropping ^{k} for simplicity) are:

$$\pi_s^{ds} = D_s(\tilde{x})\,(p_s - c_s(1-r) - f)$$
$$\pi_p^{ds} = D_p(\tilde{x})(p_p - c_p(1-r)) + D_s(\tilde{x})\,f$$

The game solves as in the case with no data sharing with only one main novelty: because sellers are more efficient on average, the platform needs to increase the fee to sustain local monopolies - i.e., it is harder, everything else being equal, to uncover the market if the rival is more efficient. In the appendix we prove that:

Proposition 13. *Assume there is data sharing. The platform set a fee*

$$f_{ds}^{\dagger} = \frac{2u - 1 + r}{4}$$

such that the equilibrium market configuration is local monopolies (lm) in every market i. The equilibrium prices are:

$$p_s^i = \frac{u + c_s^{i,CR} + f_{ds}^{\dagger}}{2} \qquad p_p^i = \frac{u + c_p^{CR}}{2};$$

Proof. See the Appendix. □

Both the seller and the platform are better off by sharing cost-reducing data. Intuitively, the sellers obtain a better technology (except for the seller producing at $c_s = 0$), which has a clear positive effect on their payoffs. The platform too benefits from sharing data. In fact, because it is always able to set a fee that ensures each market is never covered, it never suffers from competing against a more efficient rival.

Because sellers are more efficient, the platform has to charge a larger fee to ensure there is no competition between itself and the sellers.

Also, by raising the fee, the platform can increase the rent extraction from the sellers without altering too much the number of goods traded in the market. As a consequence, both the platform and the sellers are better off in aggregate.

From a social perspective, because data sharing lowers the monopoly price set by each seller in all markets, the effect on consumer surplus is positive. Together with the positive effects on aggregate profits, it implies that data sharing is welfare-improving. Not all sellers are better off, however. Those sellers that are ex-ante very efficient (c_s very low) obtain very small benefits from data sharing but have to pay a disproportionately larger fee. Indeed, the larger the cost-reducing effect of data, the more severely efficient sellers are hurt by data sharing. Similarly, if the cost-reducing potential of data is limited, also very inefficient sellers are negatively affected by data-sharing. In their case, the cost-reducing effect may be insufficient to compensate for the increase in the fee.

Their losses are compensated by the benefits of other sellers.

Proposition 14. *Data sharing benefits sellers that are ex-ante moderately efficient* ($\min\left\{\frac{4u+2-r}{8-4r}, 1\right\} > c_s^i > \frac{1}{4}$)*. Instead, data sharing always hurts ex-ante efficient sellers* ($c_s^i < \frac{1}{4}$)*. Moreover, if* $r \in \left(0, \frac{6-4u}{3}\right]$*, very inefficient sellers* ($c_s^i \in \left[\frac{4u+2-r}{8-4r}, 1\right]$) *are also hurt by data sharing.*

Proof. See the Appendix. □

5 Discussion and conclusions

Consumer data are becoming an essential input in the digital economy. In particular hybrid marketplaces can collect vast amounts of data, transform them into valuable information and then use them to compete against the sellers they host. In this work, we investigate the effect of a mandated data sharing policy on market outcomes across different data functionalities (price discrimination and cost reduction) and different market structures (perfect and imperfect competition).

The previous literature (Prüfer and Schottmüller, 2021; Krämer and Shekhar, 2022) has focused on the effects that mandated data sharing can have on platforms' incentives to innovate, highlighting how the effect on consumer surplus can be either positive or negative depending on the model's characteristics. When data enables price discrimination, allowing all firms to obtain them usually benefits consumers, as firms engage in price wars (see Montes et al. (2019); Bounie et al. (2021); Abrardi et al. (2022) among others). However, as far as we know, we are the first to analyze the effects of mandated data sharing when the platform can compete with downstream sellers.

We find that mandated data sharing has no effects on welfare if data can be used to price discriminate consumers who buy homogeneous goods. Indeed, we show that the platform sets a per-transaction fee that makes all sellers less efficient than the platform itself. Hence, with or without data, the sellers can only set a price equal to their marginal costs and cannot exploit consumers' data. The platform may also decide to operate as an intermediary in some markets, leaving some sellers without competitive pressure and free to perfectly price-discriminate consumers.

Instead, mandated sharing of cost-reducing data improves welfare by lowering the average price in the markets. Sellers are unaffected, as price competition drives profits to zero, but the platform is better off as more market coverage, and in turn revenues from intermediation, increase.

When goods are horizontally differentiated, data sharing (with price discrimination) may instead negatively affect consumers. The reason is that the platform can avoid data-induced price wars by setting a high per-transaction fee, which incentivizes sellers to set monopolistic prices, even when a market is fully covered. By avoiding downstream competition, both the sellers and the platform can use data-enabled price discrimination to extract higher surplus from consumers, leaving them worse off.

On the other hand, when data allow cost reduction, consumers are better off and total welfare increases. Interestingly, although the aggregate effect on sellers is positive, sellers are heterogeneously affected by mandated data sharing. In particular, efficient sellers have little gains from the cost-reducing data but suffer from the generalized increase in transaction fees, while inefficient sellers are not able to

compensate their high marginal costs in order to have an advantage over the platform. Those sellers which are mildly efficient are the only ones benefiting by a cost-reducing data sharing.
These results highlight the complexity of the effects of a mandated data sharing policy, as they are ambiguous and hard to predict. Indeed, we argue that in markets where competition is softer, mandatory data sharing may damage the very agents it is intended to protect, namely consumers and (efficient) sellers. Similarly, even when perfect competition is in place, mandated data sharing may push the platform to not enter the market, thus removing every competitive constraint from sellers which can then perfectly price discriminate consumers, bringing their net utility to zero.

Turning to platform profits, our analysis shows that they increase under data sharing. Then, a question naturally arises: if platforms unambiguously benefit from data sharing, why should a policymaker mandate it? We interpret this seemingly paradoxical result as the consequences of hidden costs we fail to model. Data sharing does not consist of a mere transfer of a file via email, but it entails investments in interoperability between sellers and buyers. Those costs could be non-negligible and could also entail competitive risks for the platform, as data sharing could stimulate entry by new platforms, exerting potential negative pressure on the incumbent. Indeed, the recitals of the DMA place market contestability among the important goals of the act. Further research is thus needed to better analyze these additional characteristics.

References

Abrardi, L., Cambini, C., Congiu, R., and Pino, F. (2022). User data and endogenous entry in online markets. *Available at SSRN 4256544*.

Anderson, S. and Bedre-Defolie, Ö. (2021). Hybrid platform model.

Aparicio, D., Metzman, Z., and Rigobon, R. (2021). The pricing strategies of online grocery retailers. NBER working paper n. 28639.

Bacchiega, E., Carroni, E., and Fedele, A. (2021). Monopolistic duopoly. *Mimeo*.

Beard, T. R., Kaserman, D. L., and Mayo, J. W. (2001). Regulation, vertical integration and sabotage. *The Journal of Industrial Economics*, 49(3):319–333.

Bergemann, D. and Bonatti, A. (2019). Markets for information: An introduction. *Annual Review of Economics*, 11:85–107.

Bhargava, H. K. and Choudhary, V. (2008). Research note—when is versioning optimal for information goods? *Management Science*, 54(5):1029–1035.

Bounie, D., Dubus, A., and Waelbroeck, P. (2021). Selling strategic information in digital competitive markets. *The RAND Journal of Economics*, 52(2):283–313.

Bourreau, M. and Gaudin, G. (2022). Streaming platform and strategic recommendation bias. *Journal of Economics & Management Strategy*, 31(1):25–47.

Campbell, J., Goldfarb, A., and Tucker, C. (2015). Privacy regulation and market structure. *Journal of Economics & Management Strategy*, 24(1):47–73.

Ciotti, F. and Madio, L. (2022). Competition for prominence. *Mimeo*.

De Corniere, A. and Taylor, G. (2020). Data and competition: a general framework with applications to mergers, market structure, and privacy policy.

Delbono, F., Reggiani, C., and Sandrini, L. (2021). Strategic data sales to competing firms. Technical Report JRC126568, JRC Digital Economy Working Paper, Seville, Spain.

Dryden, N., Khodjamirian, S., and Padilla, J. (2020). The simple economics of hybrid marketplaces.

Economides, N. (1998). The incentive for non-price discrimination by an input monopolist. *International Journal of Industrial Organization*, 16(3):271–284.

Etro, F. (2021a). Hybrid marketplaces with free entry of sellers. *Available at SSRN*.

Etro, F. (2021b). Product selection in online marketplaces. *Journal of Economics & Management strategy*, 30(3):614–637.

Foerderer, J., Kude, T., Mithas, S., and Heinzl, A. (2018). Does platform owner's entry crowd out innovation? evidence from google photos. *Information Systems Research*, 29(2):444–460.

Goldfarb, A. and Tucker, C. (2019). Digital economics. *Journal of Economic Literature*, 57(1):3–43.

Hagiu, A. (2009). Two-sided platforms: Product variety and pricing structures. *Journal of Economics & Management Strategy*, 18(4):1011–1043.

Kabiraj, T. and Lee, C. C. (2011). Licensing contracts in hotelling structure. *Theoretical Economics Letters*, 1(3):57–62.

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

Krämer, J. and Shekhar, S. (2022). Regulating algorithmic learning in digital platform ecosystems through data sharing and data siloing: Consequences for innovation and welfare. *Available at SSRN*.

Liu, Z., Zhang, D., and Zhang, F. (2021). Information sharing on retail platforms. *Manufacturing and Service Operations Management*, 23(3):547–730.

Madsen, E. and Vellodi, N. (2022). Insider imitation. *working paper*.

Mikians, J., Gyarmati, L., Erramilli, V., and Laoutaris, N. (2012). Detecting price and search discrimination on the internet. In *Proceedings of the 11th ACM Workshop on Hot Topics in Networks*, HotNets-XI, pages 79–84, New York, NY, USA. Association for Computing Machinery.

Montes, R., Sand-Zantman, W., and Valletti, T. (2019). The Value of Personal Information in Online Markets with Endogenous Privacy. *Management Science*, 65(3):955–1453.

Padilla, J., Perkins, J., and Piccolo, S. (2022). Self-preferencing in markets with vertically integrated gatekeeper platforms. *The Journal of Industrial Economics*, 70(2):371–395.

Pino, F. (2022). The microeconomics of data–a survey. *Journal of Industrial and Business Economics*, 49(3):635–665.

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

Raith, M. (1996). A general model of information sharing in oligopoly. *Journal of economic theory*, 71(1):260–288.

Rhodes, A. and Zhou, J. (2022). Personalized pricing and competition. *SSRN Working Paper 4103763*.

Thépot, J. (2007). Prices as strategic substitutes in a spatial oligopoly. *Available at SSRN 963411.*

Thisse, J.-F. and Vives, X. (1988). On The Strategic Choice of Spatial Price Policy. *The American Economic Review*, 78(1):122–137.

Wen, W. and Zhu, F. (2019). Threat of platform-owner entry and complementor responses: Evidence from the mobile app market. *Strategic Management Journal*, 40(1).

Zhu, F. and Liu, Q. (2018). Competing with complementors: An empirical look at amazon.com. *Strategic Management Journal*, 39(10):2618–2642.

A Appendix

A.1 Price discrimination with differentiated goods

Proof of Proposition 9. To solve the model, we first focus on a given market k. Depending on the value of f, the market can either be fully or partially covered. Moreover, the seller can choose to either set his price following standard Hotelling competition or to set it as a local monopolist. When useful, the superscript indicates the level of market coverage (hd - hotelling duopoly, dm - duopolistic monopoly lm - local monopoly respectively), while the subscript indicates the seller's pricing strategy (c - competitive and m - monopoly respectively).

First, suppose that the seller sets his price as a local monopolist and that f is so high that the market is partially covered. The platform extracts all surplus from consumers through tailored prices. FOCs of the seller profits with respect to its price leads to

$$p_{s_m}^{lm} = \frac{c_s + f + u}{2}; \qquad p_{p_m}^{lm} = u - t + tx.$$

The seller opts to only pass half of the fee to consumers, as a way to obtain a higher market share. Indifferent consumers' locations are

$$\tilde{x}_{s,no} = \frac{u - c_s - f}{2t}; \qquad \tilde{x}_{p,no} = \frac{c_p + t - u}{t},$$

while profits are

$$\pi_{s_m}^{lm} = \frac{(c_s + f - u)^2}{4t}; \qquad \pi_{p_m}^{lm} = \frac{c_p^2 - f(c_s + f) - 2c_p u + u(f + u)}{2t}.$$

These results hold as long as the seller can make positive profits and market shares do not overlap. The seller obtains positive profits as long as it can at least profitably serve the consumer located in $x = 0$. Thus, if $f > u - c_s$, the seller would not enter the market and the platform could not enter. Instead, by equating the indifferent consumers' locations, we find that if $f \leq 3u - 2t - c_s - 2c_p$, then market shares overlap.

When market shares overlap, the seller's pricing strategy does not change, as he still prices as a local monopolist. On the other hand, the platform adjusts his pricing strategy: while the platform can still extract all surplus from consumers located in $(\tilde{x}_{s,no}, 1]$, he instead must beat the seller's offer for consumers located in $[\tilde{x}_{p,no}, \tilde{x}_{s,no}]$. Thus, on this last segment, the platform sets a tailored price equal to

$$p_{p_m}^{dm} = \frac{c_s + f - 2t + u + 4tx}{2},$$

which results in the indifferent consumer being located in

$$\tilde{x}_{s,p} = \frac{2c_p - c_s - f + 2t - u}{4t}.$$

This results in profit being

$$\pi_{s_m}^{dm} = \frac{(c_s + f - u)(-2c_p + c_s + f - 2t + u)}{8t};$$

$$\pi_{p_m}^{dm} = \frac{4c_p^2 - c_s^2 - 6c_s f - 5f^2 - 4tc_s + 4tf - 4t^2 + 2u(3c_s + f + 6t) - u^2 - 4c_p(c_s - f + 2t + u)}{16t}.$$

Next, suppose instead that the seller adopts competitive pricing, and that f is low enough that we have full market coverage. In this scenario we have standard Hotelling competition, where both the seller and the platform simultaneously set their uniform prices. Then, the platform will set the tailored prices. By standard computations, we obtain equilibrium prices

$$p_{s_c}^{hd} = \frac{c_s + c_p + 2f + t}{2}; \qquad p_{p_c}^{hd} = cp + f.$$

As we can see, the platform sets his uniform price as low as possible, since in equilibrium all the consumers he serves will purchase through tailored prices, which are equal to

$$p_{p_c}^{hd}(x) = \frac{c_s + c_p + 2f - t + 4tx}{2}.$$

These prices lead to the indifferent consumer being located in

$$\tilde{x}_{s,p} = \frac{c_p - c_s + t}{4t},$$

while equilibrium profits are equal to

$$\pi_{s_c}^{hd} = \frac{(c_p - c_s + t)^2}{8t}; \qquad \pi_{p_c}^{hd} = \frac{(c_p - c_s)^2 + 2t(8f + 3(c_s - c_p)) + 9t^2}{16t}.$$

From the platform's profits function, it is clear that his profits are increasing in f. In turn, the fee is directly passed to the consumers, both by the seller and by the platform. By analysing consumer utility, we find that the consumer with the lowest utility after purchase is the one located in $x = 1$, as the platform can extract most of her surplus. Thus, these results hold as long as the net utility of the consumer located in $x = 1$ is ≥ 0, which translates to $f \leq \frac{2u - 3t - c_p - c_s}{2}$.

When $f > \frac{2u-3t-c_p-c_s}{2}$, the strategies regarding uniform pricing do not change; however, the platform must change the tailored prices he proposes to consumers located on the segment $[\frac{2u-t-2f-c_p-c_s}{2t}, 1]$ in order to allow consumers to maintain non-negative utility. To those consumers, the platform offers a tailored price equal to

$$p_{p_c}^{hd}(x) = u - t(1 - x) - c_p,$$

leading to profits equal to

$$\pi_{p_c}^{hd} = \frac{-c_p^2 - 6c_pc_s - c_s^2 - 8fc_p - 8fc_s - 8f^2 - 18tc_p - 6tc_s - 8ft - 9t^2 + 8u(c_p + c_s + 2f + 3t) - 8u^2}{16t}.$$

As f increases, consumers' net utility decreases. In particular, the next threshold is reached when the indifferent consumer $\tilde{x}_{s,p}$ net utility is equal to 0, which translates to $f = \frac{4u-3t-c_s-3c_p}{4}$. When $f \geq \frac{4u-3t-c_s-3c_p}{4}$, duopolistic monopoly ensues. While the seller loses market share due to the increasing fee, the platform can poach these consumers through tailored prices, leading to profits equal to

$$\pi_{s_c}^{dm} = \frac{(c_p - c_s + t)(2u - t - 2f - c_p - c_s)}{4t}$$

$$\pi_{p_c}^{dm} = \frac{-5c_p^2 - c_s^2 - 8fc_s - 12f^2 - 6tc_s - 16tf - 9t^2 - 2c_p(3c_s + 8f + 9t - 8u) + 8u(c_s + 3f + 3t) - 12u^2)}{8t}$$

Finally, as f increases further, the consumers lost by the seller are too far from the platform to be poached. This happens when $\tilde{x}_{s,no} = \tilde{x}_{p,no}$, which results in $f = \frac{4u-3t-c_s-3c_p}{2}$. Above this threshold, the platform and seller become local monopolists: the platform extracts all available surplus from his consumers, while the seller continues to price competitevely by construction. The seller's profits maintain the same form as above, while platform profits are

$$\pi_{p_c}^{lm} = \frac{c_p^2 - f(c_s + 2f + t) + 2fu + u^2 - c_p(f + 2u)}{2t}$$

Finally, when $f \geq \frac{4u-3t-c_s-3c_p}{2}$, the seller cannot profitably serve any consumer, and thus he does not enter the market, and neither can the platform.

Having analysed all cases, we now focus on the seller's pricing strategy as a function of f. By comparing the profits function, we find that if $f < \bar{f}$ the seller obtains higher profits with competitive pricing, while otherwise he opts for monopolistic pricing. $\bar{f}$ is found by equating the seller's profits function, and is equal to

$$\bar{f} = c_s - c_p - t + \sqrt{2c_s^2 - c_p^2 - 2tc_p - t^2 + 2c_pu - 4c_su + 2tu + u^2}$$

Thus, seller's and platform's profits as a function of f are

$$\pi_s = \begin{cases} \frac{(c_p-c_s+t)^2}{8t} \quad \text{for} \;\; 0 \leq f < \frac{4u-3t-c_s-3c_p}{4} \\ \frac{(c_p-c_s+t)(2u-t-2f-c_p-c_s)}{4t} \quad \text{for} \;\; \frac{4u-3t-c_s-3c_p}{4} \leq f < \bar{f} \\ \frac{(c_s+f-u)(-2c_p+c_s+f-2t+u)}{8t} \quad \text{for} \;\; \bar{f} \leq f < 3u - 2t - c_s - 2c_p \\ \frac{(c_s+f-u)^2}{4t} \quad \text{for} \;\; 3u - 2t - c_s - 2c_p \leq f < u - c_s \\ 0 \quad \text{for} \;\; f \geq u - c_s \end{cases}$$

$$
\pi_p = \begin{cases}
\frac{(c_p-c_s)^2+2t(8f+3(c_s-c_p))+9t^2}{16t} & \text{for} \quad 0 \leq f < \frac{2u-3t-c_p-c_s}{2} \\
\frac{-c_p^2-6c_pc_s-c_s^2-8fc_p-8fc_s-8f^2-18tc_p-6tc_s-8ft-9t^2+8u(c_p+c_s+2f+3t)-8u^2}{16t} & \text{for} \quad \frac{2u-3t-c_p-c_s}{2} \leq f < \frac{4u-3t-3c_p-c_s}{4} \\
\frac{-5c_p^2-c_s^2-8fc_s-12f^2-6tc_s-16tf-9t^2-2c_p(3c_s+8f+9t-8u)+8u(c_s+3f+3t)-12u^2)}{8t} & \text{for} \quad \leq f < \bar{f} \\
\frac{4c_p^2-c_s^2-6c_sf-5f^2-4tc_s+4tf-4t^2+2u(3c_s+f+6t)-u^2-4c_p(c_s-f+2t+u)}{16t} & \text{for} \quad \bar{f} \leq f < 3u-2t-c_s-2c_p \\
\frac{c_p^2-f(c_s+f)-2c_pu+u(f+u)}{2t} & \text{for} \quad 3u-2t-c_s-2c_p \leq f < u-c_s \\
0 \quad \text{for} \quad f \geq u-c_s
\end{cases}
$$

To find the equilibrium fee, we convert all the thresholds on f to thresholds on c_p, and maximize platform's profits across all markets:

$$
\max_f \int_0^1 \pi_p \, dc_s
$$

Standard calculations yield to $f^* = \frac{2u+4t+4c_p-3}{10}$, which corresponds to all markets being under duopolistic monopoly with the sellers pricing as a local monopolist. Replacing f^* in firms' prices and profits gives the results described in the Proposition.

Proof of Proposition 10. When both the platform and sellers have data, they price discriminate all consumers they serve. Thus, their uniform prices do not influence their strategies.

First, suppose that f is so high that the market is partially covered: then, as no consumer can be reached by both the platform and the seller, each of them will set their tailored prices to extract all surplus from each consumer. This leads to

$$
p_s^{TM}(x) = u - tx; \qquad p_p^{TM}(x) = u - t + tx.
$$

the location of the last consumer buying from the seller and the platform are respectively

$$
\tilde{x}_{s,no} = \frac{u - c_s - f}{t}; \qquad \tilde{x}_{p,no} = \frac{c_p + t - u}{t},
$$

which results in profits being

$$
\pi_s^{lm} = \frac{(c_s + f - u)^2}{t}; \qquad \pi_p^{lm} = \frac{c_p^2 - 2f(c_s + f) - 2uc_p + 2fu + u^2}{2t}.
$$

When $f \geq u - c_s$, the seller cannot profitably serve any consumer and leaves the market, thus also impeding entry to the platform. Instead, when $f < 2u - t - c_p - c_s$ we have $\tilde{x}_{s,no} > \tilde{x}_{p,no}$, and the consumers in $[\tilde{x}_{p,no}, \tilde{x}_{s,no}]$ become contestable.

Let us focus on one of these contestable consumers. If she is served by the seller, the platform obtains profits equal to f. Instead, if she is served by the platform, the platform obtains $p_p^{TM}(x) - c_p$.

By standard calculations, we find that the platform is better off by leaving to the seller all the consumers located in $[\tilde{x}_{p,no}, \tilde{x}_{p2,no}]$, where $\tilde{x}_{p2,no} = \frac{c_p+f+t-u}{t}$. Intuitevely, $\tilde{x}_{p2,no}$ is the last consumer from which the platform can extract revenues equal to $c_p + f$. Thus, it is more profitable for the platform to leave consumers in $[\tilde{x}_{p,no}, \tilde{x}_{p2,no}]$ to the seller, as he his more efficient in extracting surplus. Thus, the seller's profits are the integral of the monopolistic tailored price from 0 to $\tilde{x}_{s,no}$, while the platform's profits are the integral of his monopolistic tailored price from $\tilde{x}_{s,no}$ to 1 plus the agency profits $f\tilde{x}_{s,no}$, resulting in

$$\pi_s^{dm} = \frac{(c_s + f - u)^2}{t}$$

$$\pi_p^{dm} = \frac{-c_s^2 - 3f^2 - 2ft - t^2 - 2c_s(2f + t - 2u) - 2c_p(c_s + f + t - u) + 6fu + 4tu - 3u^2}{2t}.$$

Finally, when $\tilde{x}_{s,no} > \tilde{x}_{p2,no}$, the platform starts contesting consumers located in $[\tilde{x}_{p2,no}, \tilde{x}_{s,no}]$ as it can extract from them revenues higher than f. This scenario holds whenever $f < \frac{2u-t-c_p-c_s}{2}$. Price competition follows the following strategies:

- Consumers located in $[0, \tilde{x}_{p2,no})$ are poached by the seller through his monopolistic tailored price $p_s^{TM}(x)$;
- Consumers located in $[\tilde{x}_{p2,no}, \tilde{x}_{s,no})$ are contested;
- Consumers located in $[\tilde{x}_{s,no}, 1]$ are poached by the platform through his monopolistic tailored price $p_p^{TM}(x)$.

Let us focus on the second segment: both the platform and the seller will set their competitive tailored prices to beat the opponent's best offer, as Bertrand competition ensues. The seller will not price any lower than $c_s + f$, while the platform will not price lower than $c_p + f$, leading to

$$p_s^{TC}(x) = c_p + f + t - 2tx; \qquad p_p^{TC}(x) = c_s + f - t + 2tx.$$

The indifferent consumer in the second segment is thus located in

$$\tilde{x}_{s,p} = \frac{c_p - c_s + t}{2t},$$

and profits are equal to

$$\pi_s^{hd} = \frac{-c_p^2 + c_s^2 - 2f^2 - 2c_st - 4ft - t^2 - 2c_p(c_s + 2f + t - 2u) + 4u(f + t) - 2u^2}{4t}$$

$$\pi_p^{hd} = \frac{c_p^2 - c_s^2 - 4c_sf - 2f^2 - 2tc_s - t^2 - 2c_p(c_s + t) + 4u(c_s + f + t) - 2u^2}{4t}.$$

Platform's profits as a function of f are

$$\pi_p = \begin{cases} \frac{c_p^2 - c_s^2 - 4c_s f - 2f^2 - 2tc_s - t^2 - 2c_p(c_s+t) + 4u(c_s+f+t) - 2u^2}{4t} & \text{for } 0 \leq f < \frac{2u-t-c_p-c_s}{2} \\ \frac{-c_s^2 - 3f^2 - 2ft - t^2 - 2c_s(2f+t-2u) - 2c_p(c_s+f+t-u) + 6fu + 4tu - 3u^2}{2t} & \text{for } \frac{2u-t-c_p-c_s}{2} \leq f < 2u - t - c_p - c_s \\ \frac{c_p^2 - 2f(c_s+f) - 2uc_p + 2fu + u^2}{2t} & \text{for } 2u - t - c_p - c_s \leq f < u - c_s \\ 0 \quad \text{for } f \geq u - c_s \end{cases}$$

Having computed profits for all levels of f, we now find the equilibrium fee by maximizing platform total profits with respect to f. FOC w.r.t. to f result in $f^*_{ds} = \frac{3u-t-c_p-1}{3}$. When plugged in the functions above, we find that this fee results in all markets being under duopolistic monopoly, with both the seller and the platform only pricing through their monopolistic tailored prices.

Proof of Proposition 11. With regards to platform and sellers' profits, direct comparisons between the results described in Propositions 9 and 10 show that all firms' profits increase with data sharing. With regards to consumer surplus, recall that under no data sharing all markets are under duopolistic monopoly with the seller opting for monopolistic pricing. Consumer surplus in a given market k is thus equal to

$$CS^k = \int_0^{\tilde{x}_{p,no}} u - tx - p_{s_m}^{lm}\, dx + \int_{\tilde{x}_{p,no}}^{\tilde{x}_{s,no}} u - t(1-x) - p_{p_m}^{dm}\, dx + \int_{\tilde{x}_{s,no}}^{1} u - t(1-x) - p_{p_m}^{lm}\, dx =$$

$$= \frac{(4c_p - 3 + 10c_s + 4t - 8u)^2}{800t},$$

which, once summed across all markets leads to

$$CS = \frac{(7 + 4c_p + 4t - 8u)^3 - (4c_p - 3 + 4t - 8u)^3}{24000t}.$$

Instead, under data sharing, the seller and the platform never compete head to head for any consumer: thus, they can extract all surplus from consumers, and $CS = 0$.

By adding firms' profits and CS in the two cases, we find that total welfare increases under data sharing.

A.2 Cost reduction with differentiated goods

Proof of Proposition 12. The platform always sets a fee f such that, in equilibrium, the seller and the platform are local monopolies. To prove this result, assume the firms operate in Hotelling duopoly and the market is fully covered. In any given market k, standard calculation leads to the Hotelling prices

$$p_s^{hd} = \frac{(c_p(1-r) + 2c_s)}{3} + f + t; \qquad p_p^{hd} = \frac{(2c_p(1-r) + c_s)}{3} + f + t$$

The seller is able to pass the fee entirely on consumers, without lowering its margins. Hence, as observable from the profits, conditional on f

$$\pi_s^{hd} = \frac{(c_s - c_p(1-r) - 3t)^2}{18t}$$

$$\pi_p^{hd} = \frac{t}{2} + f - \frac{c_p(1-r)}{3} + \frac{c_s(6t - 2c_p(1-r))}{18t} + \frac{c_p^2(1-r)^2 + c_s^2}{18t}$$

the platform has strict incentives to indefinitely increase the fee. The usual result emerges, with the equilibrium fee in this specific scenario being a corner solution (i.e., $f = \bar{f} \equiv \frac{2u-3t-c_s-c_p(1-r)}{2}$ such that the indifferent consumer gets zero utility, see Figure 3).

Second, consider the intermediate case in which the two agents (platform and seller) operate in a monopolistic duopoly. Following Bacchiega et al. (2021), the two firms set the monopolistic prices:

$$p_s^{md} = u - \frac{t}{2} + \varepsilon; \qquad p_p^{md} = u - \frac{t}{2} - \varepsilon$$

with ε sufficiently small. The market is split accordingly and the profits of the two agents are:

$$\pi_s^{md} = \left(\frac{1}{2} - \frac{\varepsilon}{t}\right)\left(u - \frac{t}{2} - c_s - f + \varepsilon\right)$$

$$\pi_p^{md} = \left(\frac{\varepsilon}{t} + \frac{1}{2}\right)\left(u - \frac{t}{2} - c_p(1-r) - \varepsilon\right) + f\left(\frac{1}{2} - \frac{\varepsilon}{t}\right)$$

One can notice immediately that the platform's profits are monotonically increasing in f. Hence, conditional on consumers' willingness to pay for a good, the platform has the incentives to increase the fee indefinitely. Again, we end up with a corner solution such that $f = \bar{\bar{f}} \equiv 2(u-t) - c_s - c_p(1-r)$ (see Figure 3), i.e., the fee above which the market is partially not covered.

Finally, let us now turn to the case, where firms operate as local monopolies. In order for this scenario to exist, the consumer indifferent between purchasing the good from the seller or the platform must earn net negative utility from consumption. Hence, she prefers not consuming at all and gets zero utility. When in this scenario, both the platform and the seller set monopoly prices:

$$p_s^{lm} = \frac{c_s + f + u}{2}; \qquad p_p^{lm} = \frac{c_p(1-r) + u}{2}$$

The conditional profits level of the platform is now concave in the fee f, as one can easily observe:

$$\pi_s^{lm} = \frac{u - c_s - f}{4t}, \qquad \pi_p^{lm} = \frac{f(u-c_s) - f^2}{2t} + \frac{(u - c_p(1-r))^2}{4t}$$

Moreover, profits are bell-shaped in f, with a global maximum in $(u - c_s)/2$, which is always lower than $u - c_s$, i.e., the level above which the platform kicks the rival out of the market. The maximum can be either above $\bar{\bar{f}}$, meaning that the solution is interior, or below it, meaning that the solution is a corner. Formally $f^* = \max\{\bar{\bar{f}} + \eta;\ (u - c_s)/2\}$, with η arbitrarily small and positive, is the fee that would be chosen in that specific market.

It is easy to show that the solution is interior if:

$$u > 1 \quad \text{and} \quad t > \frac{3u - c_s - 2c_p(1-r))}{4}$$

In any case, the optimal fee lies in the region of parameters where the two economic agents (platform and seller) operate as local monopolies.

Hence, in every market i, the two goods operate in local monopolies. we use π_p^{lm} in the platform's aggregate profit function and maximize it w.r.t. f:

$$\max_f \quad \int_0^1 \pi_p^{lm}\, dc_s = \int_0^1 \frac{f(u - c_s) - f^2}{2t} + \frac{(u - c_p(1-r))^2}{4t}\, dc_s$$

Standard calculations yield to $f = f^\dagger = \frac{2u-1}{4}$. Using $f^\dagger$ in the prices of the firms in all markets allows us to derive the results in Proposition 12. □

Proof of Proposition 13. The proof unfolds as the one for Proposition 12. By the same logic as above, it is easy to show that, in each market, the profit function of the platform exhibits a maximum in the region where $f > \bar{\bar{f}}_{ds} = 2(u - t) - (c_s + c_p)(1 - r)$

The conditional profits level of the platform is now concave in the fee f, as one can easily observe:

$$\pi_{s,ds}^{lm} = \frac{u - c_s(1-r) - f}{4t}, \qquad \pi_{p,ds}^{lm} = \frac{f(u - c_s(1-r)) - f^2}{2t} + \frac{(u - c_p(1-r))^2}{4t}$$

Moreover, profits are bell-shaped in f, with a global maximum in $(u - c_s(1-r))/2$, which is always lower than $u - c_s(1 - r)$, i.e., the level above which the platform kicks the rival out of the market. The maximum can be either above $\bar{\bar{f}}_{ds}$, meaning that the solution is interior, or below it, meaning that the solution is a corner. Formally $f_{ds}^* = \max\{\bar{\bar{f}} + \eta;\ (u - c_s(1-r))/2\}$, with η arbitrarily small and positive, is the fee that would be chosen in that specific market.

It is easy to show that the solution is interior if:

$$u > 1 \quad \text{and} \quad t > \frac{3u - (c_s + 2c_p)(1-r))}{4}$$

In any case, the optimal fee lies in the region of parameters where the two economic agents (platform and seller) operate as local monopolies.

Hence, in every market i, the two goods operate in local monopolies. we use π_p^{lm} in the platform's aggregate profit function and maximize it w.r.t. f:

$$\max_{f} \quad \int_0^1 \pi_p^{lm}\, dc_s = \int_0^1 \frac{f(u - c_s(1-r)) - f^2}{2t} + \frac{(u - c_p(1-r))^2}{4t}\, dc_s$$

Standard calculations yield to $f = f_{ds}^{\dagger} = \frac{2u-1+r}{4}$. Using $f_{ds}^{\dagger}$ in the prices of the firms in all markets allows us to derive the results in Proposition 13. □

Proof of Proposition 14. Using $f^{\dagger}$ and $f_{ds}^{\dagger}$ in the profit function of sellers in each market π_s^i and $\pi_{s,ds}^i$, respectively, it is possible to derive their actual payoffs:

$$\pi_s^{i,\dagger} = \frac{(1 + 2u - 4c_s^i)^2}{64t}; \qquad \pi_{s,ds}^{i,\dagger} = \frac{(1 - r + 2u - 4c_s^i(1-r))^2}{64t};$$

Standard calculations show that:

$$\pi_s^{i,\dagger} - \pi_{s,ds}^{i,\dagger} > 0 \quad \text{if} \quad 0 < c_s^i < \frac{1}{4} \quad \text{or} \quad \frac{1}{4} \leq c_s^i < 1 \quad \text{and} \quad u < \frac{(4c_s(2-r) - 2 + r)}{4}$$

The last condition can be rearranged as $c_s > \frac{4u+2-r}{4(2-r)}$. Moreover, one can see that

$$\frac{4u + 2 - r}{4(2-r)} < 1 \quad \text{if} \quad \left(0 < u \leq \frac{3}{4} \text{ and } 0 < r < 1\right) \quad \text{or} \quad \left(\frac{3}{4} < u < \frac{3}{2} \text{ and } 0 < r < \frac{1}{3}(6 - 4u)\right)$$

Focusing on $u > 1$, the condition above reduces to

$$\frac{4u + 2 - r}{4(2-r)} < 1 \quad \text{if} \quad \left(1 < u < \frac{3}{2} \text{ and } 0 < r < \frac{1}{3}(6 - 4u)\right)$$

□