Business models, consumer data and privacy in platform markets

Jorge Padilla[1] · Salvatore Piccolo[1,2] · Helder Vasconcelos[1,3]

Abstract

We examine the drivers of the accuracy of the consumer information collected by a digital platform. In an environment where consumers mind their privacy, we compare a pure-intermediation model, where the platform plays a matching function only, by connecting buyers and sellers, with a hybrid business model, where the platform also introduces its private label to compete with third-party sellers. We show that the platform's incentive to collect demand information in the two models depends on the intensity of intra-platform competition and on its bargaining power vis-à-vis third-party sellers. When end-users perceive the platform's private label and the third-party sellers' products as relatively close substitutes (strong intra-platform competition) and the intermediary has a strong bargaining position in the negotiation with the sellers, it tends to acquire less accurate information under the hybrid model than in the pure-intermediation model, at the benefit of consumer privacy. Otherwise, more information is acquired under the hybrid model. These results shed new light on the link between alternative business models, consumer privacy and information collection in the digital sector, and may help explaining why some platforms tend to protect more consumer privacy than others.

✉ Salvatore Piccolo
  salvapiccolo@gmail.com

1 Compass Lexecon, Madrid, Spain

2 Bergamo University, Compass Lexecon and CSEF, Bergamo, Italy

3 Porto University and Compass Lexecon, Madrid, Spain

 Springer

JEL Classification D47 · D85 · L5 · L81 · M3

1 Introduction

Online platforms are the architects of the digital revolution. Thanks to these platforms, nowadays, consumers and sellers enjoy multiple trading solutions. In addition to meeting physically in stores, they can also trade in a virtual, impersonal, and presumably anonymous world. The reduction of search costs, the increased delivery speed, and higher market transparency are the bright side of this revolution. However, these companies also manage a massive amount of consumer data. Platforms such as Amazon, Apple, and Google, to name a few, collect, package, and disclose users' data to third parties that use this knowledge for commercial, marketing, and, in the worst case, fraudulent purposes. The information that platforms collect covers a broad spectrum of individual data, ranging from users' individual characteristics, such as gender, age, and location, to their browsing patterns, prior transactions, social interaction, etc. Platforms can, therefore, forecast consumers' tastes, habits, and social preferences, and monetize this information through personalized offers.[1] Consumer data may also land in wrong 'hands' and be used for illegal purposes that damage consumers and their privacy (e.g., credit card and/or identity cloning). This is allegedly the major dark side of the digital revolution.

A flourishing academic literature has started to investigate the interaction between data management, marketing strategies, and competition in platform markets (see, e.g., Bergemann & Bonatti, 2019; Jullien, 2012; Peitz & Reisinger, 2015, for recent surveys). However, these models are silent on the link between platforms' business models and the accuracy of the consumer data that they collect and eventually disclose to self-interested third-party sellers.

Notably, while some online businesses have mainly maintained a brokerage activity (e.g., eBay and Google) others operate under hybrid business models and have developed their own private labels to compete with third-party sellers operating through their marketplaces (e.g., Amazon and Apple). Do all these online intermediaries have the same incentives to acquire and disclose users' personal information? If not, what are the determinants of different approaches to information and privacy management? Is the choice of business model—i.e., pure intermediation vs. hybrid platforms one of these key factors?

In this paper, we study the drivers of the accuracy of the information that digital intermediaries collect and disclose. Specifically, we compare the incentives to collect demand information by a an online intermediary (platform) operating under two alternative business models: a pure-intermediation model, where it plays a matching function only by connecting a third-party seller with buyers, and a hybrid model where, in addition to its traditional middlemen role, the platform also introduces its own private label in the marketplace and competes with the third-party seller

---

[1] The worldwide digital advertising expenditure has been estimated at around 283 billion U.S. dollars in 2018, and is expected to grow up to 517 billion by 2023. See, e.g., https://www.statista.com/outlook/216/100/digital-advertising/worldwide

distributing through its marketplace. We argue that there is no objective presumption that pure intermediation platforms collect more or less information than hybrid platforms. Platforms' business model is not neutral to data collection. This observation should be considered carefully by privacy authorities, especially because in the EU the GDPR is based on the 'data minimization principle.'[2]

We set up a simple duopoly model with linear demand and random willingness to pay (demand intercept). The platform collects information on the demand's random component and, coherently with recent regulatory trends imposing big tech companies transparency requirements to promote level playing field competition, is mandated to disclose this information to the sellers distributing through its marketplace (including its retail unit when present). Sellers use such information to target quality (or advertising) and prices. The crucial, and somewhat novel, assumption is that disclosing more accurate information directly harms consumers because they mind their privacy.[3] To isolate the effects of contractual frictions (e.g., double marginalization) on the platform's information acquisition problem, in the baseline model we assume efficient contracting—i.e., the platform extracts a fixed share of the seller's profit.[4]

Within this setting, we show that the platform's incentive to gather information in the two business models depends on the degree of substitutability between the private label (which is present only under the hybrid model) and the product of the seller distributing through it (intra-platform competition), and the distribution of the bargaining power in the negotiation with the seller. When intra-platform competition is sufficiently intense, and the platform has a strong bargaining position in the negotiation with the seller, it tends to acquire and disclose more accurate information in the pure-intermediation model than in the hybrid model, at the expense of consumers' privacy. Otherwise, the platform collects and discloses more information when it operates a hybrid model.

Gathering information has two main effects on the platform's profit. On the one hand, regardless of its business model, the platform is willing to gather information because it allows the sellers active in its marketplace to make more accurate pricing and quality/advertising decisions, generating an extra profit that the platform (partially) extracts at the negotiation stage. On the other hand, since consumers have privacy concerns, gathering information reduces demand because (ceteris paribus) fewer consumers join the platform when they fear that online purchases endanger their privacy.

---

[2] The principle of 'data minimization' means that a data controller should limit the collection of personal information to what is directly relevant and necessary to accomplish a specified purpose. They should also retain the data only for as long as is necessary to fulfil that purpose. See, e.g., https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679

[3] This concern arises when individual level information lands through the web into the wrong hands and is misused at their expense. As highlighted, in a relatively recent speech, by the EU Competition Commissioner Margrethe Vestager: " People understand that handing over data has a cost. Because each time we share our data, we give up something very valuable. Something that could be used against us." ("When technology serves people", Brain Bar Budapest, 1 June 2018).

[4] In the Extensions, which are discussed below, we also consider unit and ad-valorem fees.

The first effect described above is present under both business models. The second effect points to less information collection in the hybrid platform than in the pure intermediation model when the private label and the product of the third-party seller are close substitutes and when the bargaining power of the platform is high. This is because when those conditions hold, the hybrid platform can appropriate (via its retailing activity) a larger share of the incremental value generated by each consumer that joins the platform and, therefore, it has a larger incentive to increase demand by protecting privacy. Hence, when the intermediary is in a strong bargaining position and intra-platform competition is strong, hybrid platforms will collect and disclose less information than pure intermediation platforms. The opposite holds otherwise.

Building on these insights, we then explore the determinants of the optimal business model. To begin with, we show that, when both models imply the same level of information accuracy, the platform prefers to operate as a pure intermediary when consumers perceive products as relatively close substitutes and its bargaining power is relatively strong. The intuition is as follows. For given information accuracy, introducing a private label in the marketplace is not worthwhile for the platform if products are close substitutes because competition erodes both the profit earned through its private label and those earned through the intermediation channel. This effect becomes even stronger, in relative terms, when the platform's bargaining strength rises, because in the pure intermediation model, the platform extracts a relatively higher share of the seller's profit. Yet, when the information policy differs in the two business models, the result is ambiguous. We find interesting cases in which the hybrid regime maximizes the platform's profit and vice-versa. The driving forces are again intra-platform competition, which makes the hybrid model relatively less appealing, and the strength of the bargaining position of the third-party seller(s), which makes it more appealing.

These findings generalize to a number of extensions which include alternative demand functions, multiple sellers and decentralized decision making within the platform. Notably, with inefficient contracting (e.g., with ad-valorem and linear per-unit fees) we find that the hybrid model always provides greater incentives to gather information under the hypothesis that the platform makes take-it-or-leave-it offers. While under ad-valorem fees this incentive unambiguously falls with competition, under linear per-unit fees we find a U-shaped relationship between the incentive to collect information and the degree of intra-platform competition. Finally, we show that the effect of the business model on consumer surplus is ambiguous, although positively biased towards the hybrid model in several cases.

2 Related literature

Our work is related to the growing theoretical literature on the economics of privacy and data collection (see Acquisti et al., 2016 for a comprehensive survey of research works on the topic). Especially relevant to our analysis is the work of Evans (2009), who shows that consumers may resist having advertising platforms collecting detailed data about their behavior and government regulation may be called for limiting the

ability of advertising intermediaries to collect these data.[5] Of particular interest is also the work by Kalvenes and Basu (2006) arguing that hybrid platforms may not want to use their private information to gain a competitive advantage over rivals to stimulate participation. Madsen and Vellodi (2021); Lam and Liu (2020), instead, study the entry decision of a platform's owner in the product space of the complementors ('hybrid business model') and the role of individualized information. In contrast to us, both these papers focus on a data dimension that affects sellers and not consumers.

The paper is also related to the extensive literature on multi-sided platforms (see, e.g., Jullien, 2012). In this literature, some authors have analyzed the trade-off between the platform business model and more traditional alternatives. In particular, Hagiu and Wright (2015) analyze the trade-offs that drive an intermediary when choosing between operating as a marketplace, as a reseller or as a hybrid. Along similar lines, Abhishek et al. (2016) identify the conditions under which agency selling (a selling format wherein online retailers allow manufacturers direct access to their consumers) should be preferred to reselling (wholesale pricing), as well as the implications of that choice for market participants (i.e., for competing on-line retailers, manufacturers and consumers). Karle et al. (2020) have more recently argued that the competitive conditions among sellers shape the market structure in platform industries—i.e., agglomeration versus segmentation. Hagiu et al. (2020) study the welfare effects of several policies that may restrict how intermediaries operate, with an emphasis on hybrid intermediaries. In their model third-party sellers also have their own store (direct distribution), where they can set prices in a differentiated way, and have a fraction of captive (single-homing) consumers that will never buy in the marketplace. However, in contrast to our model, in which quality (advertising) is endogenous, they assume that third-party sellers sell a superior product but they produce it at a higher cost than the intermediary. More recently, Anderson and Bedre-Defolie (2021) have identified the negative effects of a hybrid business model for consumers abstracting from privacy related issues. Shopova (2021), instead, studies the incentives of a hybrid platform to offer a low-quality product abstracting, however, from the implications that this strategy has on information and privacy management. None of these papers explores, however, the incentives to collect demand information of a platform operating under alternative business models in an environment where consumers mind their privacy.

3 The baseline model

Players and environment This section builds a simple environment to develop and compare two alternative digital business models: a pure- and a hybrid-intermediation model. In both models, there is a platform (P) through which a single seller (S) distributes its product (in the extensions we extend the analysis to $N > 1$ sellers). In

---

[5] As (Evans, 2009, p. 58) highlights, (overly) lenient privacy regulations can harm consumers as they " could incur the costs of having private information disclosed and potentially misused, and incur the costs of reducing their use of the web because of concerns over privacy." Further, " regardless of whether their private information is disclosed, consumers may dislike receiving ads that reflect too much knowledge about them " .

the pure intermediation model, P only plays a matching function, by connecting the two sides of the market; in the hybrid model, P also develops a private label to compete with S to serve consumers—i.e., there is a duopoly with differentiated products in the final market.

Consumers, information structure and privacy concerns There is a representative consumer who has preferences described by the following utility function

$$
U(\cdot) \triangleq \underbrace{\sum_{i=1}^{N} \left(A + \theta + x_i\right)q_i - \frac{1}{2}\sum_{i=1}^{N} q_i^2 - dq_1q_2 - \sum_{i=1}^{N} p_iq_i -}_{\text{Standard Singh-Vives Utility}} \underbrace{\eta\psi\sum_{i=1}^{N} q_i}_{\text{Privacy Concerns}}, \tag{1}
$$

where $N = 1$ in the pure intermediation model, and $N = 2$ in the hybrid model. As a convention, we assume that S supplies product 1 irrespective of the business model, while P supplies product 2 when they compete. Hence, in the above expression q_1 denotes S's output while q_2 denotes P's output.

The utility function described above can be decomposed in two parts. The first corresponds to the standard (Singh & Vives, 1984) quadratic utility function, which we have chosen for its analytical tractability (i.e., it yields linear demand functions and a tractable expression for consumer surplus). The parameter $d \geq 0$ captures, therefore, the exogenous degree of differentiation between products: the larger d, the closer substitutes products are perceived by consumers. The parameter $A > 0$ is an exogenous and deterministic component of the representative consumer's willingness to pay (the demand intercept); θ is a zero-mean random variable distributed uniformly on the support $\Theta \triangleq [-\sigma, \sigma]$, with σ capturing the heterogeneity of tastes; $x_i \geq 0$ is an endogenous (consumer-surplus enhancing) variable that we interpret as the (observable) quality of each product traded on the platform (e.g., supply of non-market activities such as product description, guarantees, post-sale services, etc.). Alternatively, one can think of x_i as being sellers' advertising intensity—i.e., the extent to which consumers are exposed to ads promoting product i.[6] Players are uninformed about θ, but P can gather an informative signal. The information policy is '*all-or-nothing*' (see, e.g., Vives, 1984; Gal-Or, 1985): P receives a signal $s \in \Theta$ which is fully informative of the state θ with probability $\eta \in [0, 1]$, and uninformative otherwise—i.e., with probability $1 - \eta$ the state of nature θ and the signal s are identically and independently distributed.

The second part of the utility function determines consumers' privacy concerns and represents a novel aspect of our analysis. Following Gal-Or et al. (2018) we assume that tracking and collecting data about users can rise concerns among them. First, although users may benefit from a targeted choice of quality or informative advertising, the resulting loss of privacy may discourage users from adopting the platform (see, e.g., Xu et al., 2012; Tucker, 2014, for empirical studies on targeting and privacy concerns). Second, this information may land in wrong 'hands' and be used for illegal purposes that damage consumers (e.g., credit card and/or identity cloning). Hence, we posit that privacy concerns are proportional to the accuracy of

---

[6] Another interpretation for x_i is targeted advertising.

the information gathered by P—i.e., the more accurate the information gathered by P, the greater the potential damage consumers face—and proportional to the amount of transactions made by consumers—i.e., the more transactions consumers make, the greater the risk that the information left on the marketplace is misused.[7] The parameter $\psi \geq 0$ measures the extent of this damage.[8]

The precision η of the signal is endogenous and chosen by P at the outset of the game. For simplicity, the cost of gathering information with precision η is linear, and denoted by $c\eta$.[9] Furthermore, we assume that regulatory constraints, imposing non-discriminatory information rules, prevent P from hiding information to S. This hypothesis is in line with many regulatory initiatives aimed at promoting level-playing field competition in digital industry. For example, the CMA has recently created a digital markets unit with the open objective of enforcing a code of conduct based on fair trading, trust, and transparency so to prevent tech giants from discriminating the third-party sellers operating on their platforms/marketplaces at the advantage of their retail harms (e.g., Amazon Retail).[10] As a result, we assume that P and S must have the same information—i.e., P is mandated to fully disclose s regardless of the business model.[11]

Contracting To focus only on the information management problem faced by the platform, in the baseline model we rule the standard double marginalization arising with per-unit and ad-valorem fees, and simply assume that P appropriates a fraction $b \in [0, 1]$ of S's profit—i.e., P's negotiation power vis-à-vis S. This assumption reflects the idea of ex-ante contracting and/or efficient bargaining, and it guarantees that when $b = 1$ the platform fully extracts S's profit.[12] In Sect. 3 (Extensions) we discuss how introducing explicit contractual frictions will alter the equilibrium of the game.

Technology Sellers have linear production technologies with marginal costs of production set to zero. This assumption is without loss of generality given the profit-sharing rule assumed above, but (as we shall discuss) it may not be innocuous when introducing ad-valorem fees.

Providing quality (advertising) is costly and, for tractability, we assume a quadratic cost function $x_i^2/2$.

---

[7] For example because the more transactions a consumer makes, the higher is the likelihood that his/her personal data are stolen or covertly sold to self-interested third parties.

[8] See, e.g., Turow et al. (2009) and Goldfarb and Tucker (2011) among others for empirical studies on these costs.

[9] Results do not change if we assume a quadratic cost of gathering information—e.g., $\frac{c\eta}{1-\eta}$.

[10] See, e.g., https://www.theguardian.com/business/2021/feb/22/uk-competition-big-tech-google-facebook-amazon-microsoft-apple-cma

[11] Padilla et al. (2021) propose a model in which even if the platform can hide information to the sellers, in most cases it wants to share this information.

[12] Formally, this means that there is no participation constraint to be checked at $b = 1$. The participation contrarian of the seller would instead be binding when they share (or bargain) over S's revenues rather than profits.

Timing and equilibrium concept Following the literature on markets for data (e.g., Bergemann & Bonatti, 2018; Bergemann & Bonatti, 2019; Kastl et al., 2018), we assume that P can commit to an information policy (accuracy) η.[13]

The timing of the game is as follows:

- P chooses η—i.e., it commits to deliver some data and related analytics with a certain quality.
- θ realizes and P observes signal s, which is then disclosed to S.
- Depending on the business model, sellers simultaneously choose qualities and set prices.
- Demand allocates, profits materialize and payments between S and P are made.

The equilibrium concept is Subgame Perfect Nash Equilibrium (SPNE).

Assumptions We assume that $A > \psi + \sigma$ to guarantee positive demand in every state. In addition, we also assume that $d \leq \frac{1}{2}$—i.e., competition is not too intense—to guarantee that in the hybrid model prices and qualities are non-negative for every level of $b \in [0, 1]$.

3.1 Pure intermediation

Consider first the pure intermediation model—i.e., $q_2 = 0$. The utility function of the representative consumer is

$$u(\cdot) = \left(A + \theta + x_1\right)q_1 - \frac{q_1^2}{2} - p_1 q_1 - \psi \eta q_1.$$

Optimizing with respect to q_1, we obtain the following linear demand function conditional on θ

$$Q_1\left(p_1, x_1, \theta\right) = A + \theta + x_1 - \psi \eta - p_1.$$

Hence, for given signal s disclosed by the platform, the seller solves the following maximization problem

$$\max_{x_1 \geq 0, p_1 \geq 0} \mathbb{E}\left[Q_1\left(p_1, x_1, \theta\right)|s\right]p_1 - \frac{x_1^2}{2},$$

where, given the all-or-nothing information structure postulated above, the conditional expectation of θ given s is

$$\mathbb{E}[\theta|s] = \eta s.$$

Maximizing with respect to p_1 and x_1, respectively, we have

---

[13] In this literature, full commitment is often justified as a reduced form of the enforcement of a given information policy through repeated interactions between the information provider and third-party sellers—e.g., third-party sellers may credibly threaten to leave the platform if they discover an accuracy lower than what expected.

$$A + \mathbb{E}[\theta|s] + x_1 - \eta\psi - 2p_1 = 0, \tag{2}$$

and

$$p_1 - x_1 = 0. \tag{3}$$

Since consumers value positively quality (advertising), products featuring higher quality (or that are advertised more intensively) will be more expensive—i.e., as x_1 grows large, p_1 increases too. Moreover, as privacy concerns become more relevant—i.e., as ψ grows large—the consumers' willingness to pay drops, which implies a lower price and thus lower quality (advertising).

Solving (2) and (3) simultaneously, it is easy to find the optimal price and the optimal targeting chosen by the seller in the pure-intermediation model—i.e.,

$$p_1^*(s) = x_1^*(s) \triangleq A + (s - \psi)\eta,$$

with $p_1^*(s) \geq 0$ since we assumed $A > \sigma + \psi$.

As intuition suggests, $p_1^*(s)$ and $x_1^*(s)$ are increasing in s—i.e., the higher the signal, the higher demand (in expected terms), and thus the higher the (equilibrium) price and quality. The effect of η is ambiguous and depends on the difference $s - \psi$ which reflects the trade-off between the demand enhancing effect and the privacy concerns. When $\psi > s > 0$, privacy concerns are relatively important, hence the price is decreasing in the precision of the information η. An interesting aspect to highlight is that quality (advertising intensity) is declining in ψ, which suggests that if a platform has a relatively poor reputation for respecting and protecting consumers' privacy, in equilibrium, the products traded on that platform are of relatively low quality or are advertised less intensively.

Substituting $p_1^*(\cdot)$ and $x_1^*(\cdot)$ into S's (gross) expected (i.e., signal contingent) profit we have

$$\pi^*(s) \triangleq \frac{[A + (s - \psi)\eta]^2}{2}.$$

The quadratic structure of the profit function is standard in oligopoly games with linear demand (see, e.g., Vives, 2010). Hence, $\pi^*(s)$ is convex in s—i.e., firms are risk-lover and their profits increase with the volatility of the signal s. The more volatile the signal, the higher the profit shared by the platform and the seller. Integrating with respect to s, the platform's expected profit net of the cost of acquiring information is

$$\pi^*(\eta) \triangleq b \int_{-\sigma}^{\sigma} \pi^*(s) \frac{ds}{2\sigma} - c\eta = b \frac{\eta^2 \left(3\psi^2 + \sigma^2\right) + 3A(A - 2\psi\eta)}{6} - c\eta, \tag{4}$$

which is clearly convex in η—i.e., P either prefers maximal noise ($\eta = 0$) or full accuracy ($\eta = 1$). Hence, P's problem

$$\max_{\eta \in [0,1]} \pi^*(\eta),$$

always features a corner solution—i.e.,

$$\eta^* \triangleq \begin{cases} 1 & \Leftrightarrow \pi^*(1) > \pi^*(0) \\ \in [0,1] & \Leftrightarrow \pi^*(1) = \pi^*(0) \\ 0 & \Leftrightarrow \pi^*(1) < \pi^*(0) \end{cases}.$$

Comparing

$$\pi^*(1) = \frac{b\left(3\psi^2 + \sigma^2 + 3A(A - 2\psi)\right)}{6} - c,$$

with

$$\pi^*(0) = b\frac{A^2}{2},$$

we can show the following:

Proposition 1 *In the pure intermediation model, P chooses $\eta^* = 1$ only if*

$$\psi \leq \psi^* \triangleq A - \sqrt{A^2 - \frac{\sigma^2}{3}} > 0,$$

and if

$$c \leq c^* \triangleq \frac{b\left(\sigma^2 - 3\psi(2A - \psi)\right)}{6}.$$

Otherwise, $\eta^* = 0$.

In addition to the fact that gathering information is costly, there are three additional effects that shape the accuracy of the information collected by P. First, since quality/advertising is costly, gathering more precise information allows P to choose to invest in quality (advertising) more efficiently—i.e., invest in quality when θ is high. Second, higher accuracy triggers privacy concerns: consumers reduce demand being concerned with their privacy. Third, since the profit function is convex in s, a higher η also benefits P because it makes its profits relatively more responsive to s, and thus more volatile.

On the net, the second effect dominates when the privacy concerns are sufficiently strong ($\psi \geq \psi^*$), in which case the platform's profit is maximized by an uninformative information policy. By contrast, when privacy concerns are not too important ($\psi < \psi^*$) gathering information with maximal accuracy is optimal provided that its cost c is not too large—i.e., $c \leq c^*$. Clearly, the convexity of the profit function with respect to s implies that as demand uncertainty increases (i.e., as σ grows large) the thresholds c^* and ψ^* rise too, so that acquiring information becomes relatively more valuable. Finally, as intuition suggests, c^* is increasing in b, meaning that P has a greater

incentive to collect information the larger is the share $b \in [0, 1]$ of S's profit that it can extract.

3.2 Hybrid platform

Consider now a hybrid platform where, in addition to allowing S to match with users, P also competes with it by developing its private label. To simplify, we begin by treating the platform and its (vertically integrated) retailing unit as a unique entity (in the extensions we show that results hold true when this assumption is relaxed).

S's and P's demand functions are, respectively,

$$q_1\left(p_1, p_2, x_1, x_2, \theta\right) \triangleq \frac{A + \theta - \psi\eta}{1 + d} + \frac{x_1 - p_1 + d\left(p_2 - x_2\right)}{1 - d^2},$$

$$q_2\left(p_2, p_1, x_2, x_1, \theta\right) \triangleq \frac{A + \theta - \psi\eta}{1 + d} + \frac{x_2 - p_2 + d\left(p_1 - x_1\right)}{1 - d^2}.$$

Hence, conditional on observing signal s, P solves the following maximization problem

$$\max_{x_2 \geq 0, p_2 \geq 0} \mathbb{E}\left[q_2(\cdot)|s\right]p_2 - \frac{x_2^2}{2} + b\left(\mathbb{E}\left[q_1(\cdot)|s\right]p_1 - \frac{x_1^2}{2}\right),$$

whose first-order conditions with respect to p_2 and x_2 are, respectively,

$$\underbrace{\frac{A + (s - \psi)\eta}{1 + d} + \frac{x_2 - dx_1 - 2p_2 + dp_1}{1 - d^2}}_{\text{Standard price-setting rule}} + \underbrace{\frac{bd}{1 - d^2}p_1}_{\text{Cross-demand effect}} = 0, \tag{5}$$

and

$$\underbrace{\frac{1}{1 - d^2}p_2}_{\text{Demand-enhancing effect}} - \underbrace{x_2}_{\text{Cost of quality}} - \underbrace{\frac{bd}{1 - d^2}p_1}_{\text{Cross-demand effect}} = 0. \tag{6}$$

The first-order condition (5) reflects three intuitive economic forces. First, when P increases the price p_2, it earns a higher profit on the infra-marginal units (i.e., on the consumers that keep purchasing the product even if its price has slightly increased). Second, a higher p_2 also reduces demand for P's product, which means a lower sales volume. Third, since prices are strategic complements, a higher p_2 also boosts S's profit which, in turn, benefits P because it extracts the share $b \in [0, 1]$ of such profit. Notice that when S increases its quality—i.e., when x_1 increases—P has a weaker incentive to increase its price because the demand for its brand drops further.

The first two terms in condition (6) reflect the same trade-off discussed in the pure intermediation model. The only difference being that the marginal benefit of investing more in quality or advertising (the first term in Eq. 6) now depends, and is increasing, in d—i.e., the closer substitutes products are, the more quality

becomes a competitive instrument to gain market share. However, since P partly internalizes S's profit, a higher x_2 also hurts P because it reduces S's demand, and therefore the profit that they share.

S solves the following maximization problem

$$\max_{x_1 \geq 0, p_1 \geq 0} \mathbb{E}\big[q_1(\cdot)|s\big]p_1 - \frac{x_1^2}{2}.$$

The first-order conditions with respect to p_1 and x_1 are, respectively,

$$\underbrace{\frac{A + \eta(s - \psi)}{1 + d} + \frac{x_1 - dx_2 - 2p_1 + dp_2}{1 - d^2}}_{\text{Standard price setting rule}} = 0, \tag{7}$$

and

$$\underbrace{\frac{p_1}{1 - d^2}}_{\text{Demand-enhancing effect}} - \underbrace{x_1}_{\text{Cost of quality}} = 0. \tag{8}$$

The intuition behind these expressions is the same as in the pure-intermediation model. Solving the system of first-order conditions (5)–(8) we can state the following result.

Proposition 2 *In the hybrid model, the equilibrium market outcome is such that*

$$p_2^\star(s) \triangleq \frac{\left(1 - d^2\right)\left(1 - d(1 + d) - d^2 b\right)}{1 - bd^2\left(1 - d^2\right) - d^2\left(3 - d^2\right)}(A + \eta(s - \psi))$$

$$\leq p_1^\star(s) \triangleq \frac{\left(1 - d^2\right)(1 - d(1 + d))}{1 - bd^2\left(1 - d^2\right) - d^2\left(3 - d^2\right)}(A + \eta(s - \psi)),$$

with equality only at $b = 0$, and

$$x_2^\star(s) = \frac{1}{1 - d^2}\big[p_2^\star(s) - bdp_1^\star(s)\big] \leq x_1^\star(s) = \frac{p_1^\star(s)}{1 - d^2},$$

with equality holding again at $b = 0$ only.

Interestingly, P's private label is cheaper and features lower quality than S's product. The reason is straightforward: P has a weaker incentive to invest in quality (or advertising) than S because it internalizes the negative effect of a higher x_2 on S's demand. Essentially, P purposefully creates a sort of 'vertical differentiation' between its product and S's product by reducing x_2, so as to relax competition and increase profits. Clearly, since P invests in quality (or advertising) less than S, product 2 must also be cheaper than product 1.

We can now turn to study the optimal accuracy of the information collected by P.
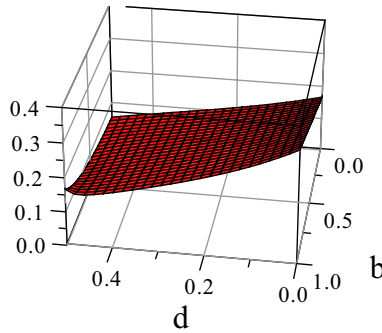
Fig. 1 Impact of b and d on profits

Lemma 3 *There exists a positive function $\Gamma(b, d)$ such that, conditional on observing signal s, P's expected profit is*

$$\pi^\star(s) \triangleq [A + \eta(s - \psi)]^2 \Gamma(b, d).$$

Hence, P's ex-ante expected profit, net of the cost of collecting information with accuracy η is

$$\pi^\star(\eta) \triangleq \left(\eta^2\left(3\psi^2 + \sigma^2\right) + 3A(A - 2\psi\eta)\right)\Gamma(b, d) - c\eta. \tag{9}$$

Once again, the expected profit features a quadratic structure and is convex in η – i.e., P's problem features corner solutions

$$\eta^\star \triangleq \begin{cases} 1 & \Leftrightarrow \pi^\star(1) > \pi^\star(0) \\ \in [0, 1] & \Leftrightarrow \pi^\star(1) = \pi^\star(0) \\ 0 & \Leftrightarrow \pi^\star(1) < \pi^\star(0) \end{cases}.$$

The function $\Gamma(b, d) \geq 0$ is characterized in the Appendix, and is plotted in Fig. 1 (red curve) below in the space $(b, d) \in [0, 1] \times [0, 1/2)$.

It can be seen that $\Gamma(b, d)$ is increasing in b (for given d) and decreasing in d (for given b). Hence, for given information disclosure, P's expected profit is decreasing with the intensity of competition and increasing with its bargaining power vis-à-vis S.

Comparing

$$\pi^\star(1) = \left(3\psi^2 + \sigma^2 + 3A(A - 2\psi)\right)\Gamma(b, d) - c,$$

with

$$\pi^\star(0) = A^2\Gamma(b, d),$$

we can state the following.

Proposition 4 *In the hybrid model, the optimal accuracy is such that $\eta^\star = 1$ only if $\psi \leq \psi^*$ and if*

$$c \leq c^\star \triangleq \left(\sigma^2 - 3\psi(2A - \psi)\right)\Gamma(b,d).$$

Otherwise, $\eta^\star = 0$.

The intuition is the same as in the pure intermediation model. As long as privacy concerns are not too strong and gathering information is not too costly, P becomes fully informed. Otherwise, it has no incentive to collect information. Of course, this incentive now depends on the extent to which consumers perceive products as more or less substitutes, and on the share b of S's profit extracted by P. As noted before, Fig. 1 shows that the incentive to become informed in the hybrid model is decreasing in d and increasing in b.

3.3 The drivers of information accuracy

We can finally study the drivers of the information accuracy in the two business models. As a benchmark, it is useful to start by considering two extreme but interesting scenarios: $b = 0$ so that P does not internalize S 's profit at all—e.g., because there is a competing platform that stands ready to attract S—and $b = 1$ where P fully internalizes S's profit—e.g., because P is monopolist in the platform market and S has no other distribution options.

Comparing the thresholds $c^\star$ and c^*, we have the following preliminary lemma:

Lemma 5 *Suppose that $\psi \geq \psi^*$, then P does not collect information regardless of the business model. By contrast, for $\psi < \psi^*$, the following holds*:

- For $b \to 0$ the incentive to collect information is stronger in the hybrid model than in the pure intermediation model for every admissible level of d—i.e., $c^\star > c^*$ for every $d \in [0, \frac{1}{2})$.
- For $b \to 1$ there exists a value $d^* \approx 0.4$ such that the incentive to acquire information is stronger in the pure intermediation model than in the hybrid model if and only if $d \geq d^*$, and the opposite is true otherwise—i.e., $c^\star \leq c^*$ iff $d \geq d^*$.

This result shows that the relationship between the incentive to collect consumer data and the business model is non-monotonic, and depends both on P's bargaining power in the negotiation with S and on the intensity of competition.

When b is sufficiently small, P gains more from collecting information in the hybrid model than in the pure intermediation model. The intuition is that when P acts as a pure intermediary, its main source of profit is the surplus extracted from S. Clearly, if the share of S's profit internalized by P is small enough ($b \to 0$), paying the cost of collecting information is not worthwhile. By contrast, since under the hybrid model, P sells its own product, it has a stronger incentive to
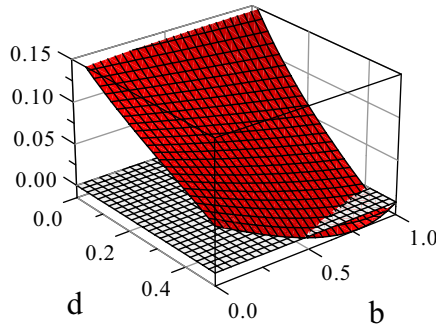
Fig. 2 Incentives to invest and business model

collect information even when $b \rightarrow 0$: an informed P can price and target quality/advertising more accurately.

By contrast, when b is sufficiently large, the incentives to collect information in the two business models may reverse depending on the intensity of intra-platform competition, as reflected by the magnitude of the differentiation parameter d. The platform has a stronger incentive to collect information in the pure intermediation model compared to the hybrid model when intra-platform competition is sufficiently intense. To see why, consider the extreme case in which $b = 1$—i.e., P fully extracts S's profit under both models. In the pure-intermediation model, S acts as a monopolist, and collecting information has the exclusive benefit of enabling it to price and invest in quality (advertising) more accurately. The implied extra profit is then fully extracted by P. In the hybrid model, instead, P and S compete fiercely when d is sufficiently large since consumers perceive their goods as relatively close substitutes. In this case, P has a weaker incentive to gather information because when S is uninformed it makes less accurate quality/advertising and price decisions, whereby competing less aggressively with P.

We can now turn study the difference between $c^\star$ and c^* for every admissible $b \in (0, 1)$. It is easy to show that

$$c^\star \gtrless c^* \quad \Leftrightarrow \quad \Gamma(b, d) \gtrless \frac{b}{6}.$$

Figure 2 below plots the difference $\Gamma_h(b, d) - \frac{b}{6}$ (red curve) in the relevant range of parameters.

The figure shows that the difference $c^\star - c^*$ is negative when b and d are both sufficiently large. Hence, P has a weaker incentive to collect information in the hybrid model than in the pure intermediation model when intra-platform competition is relatively intense (i.e., when d is sufficiently large) and when the platform has strong bargaining position vis-à-vis the sellers operating through it (i.e., when b is sufficiently large). The following then holds.

Proposition 6 *There exists a $d_0 < \frac{1}{2}$ such that*:

- for every $d \leq d_0$, P has a stronger incentive to collect information in the hybrid model for every $b \in [0, 1]$.
- for every $d > d_0$, there exists a function $b_0(d) < 1$, with $b_0'(d) < 0$, such that P has a stronger incentive to collect information in the pure intermediation model if and only if $b \geq b_0(d)$.

The intuition follows from the above discussion.

3.4 Comparison of business models

We now compare the equilibrium outcome across the two regimes. We start by considering how prices and targeting change across the two business models.

- In the hybrid model, regardless of whether P collects information or not, it always charges a price lower than that charged by S in the pure intermediation model. The same holds for the quality choice.

The reason is that in the hybrid model there is intra-platform competition, whereas in the pure intermediation model S is a monopolist.

- In the hybrid model, regardless of whether P collects information or not, S sets a lower price than in the pure intermediation model for low values of d and b. The same holds for the targeting choice.

The reason is that as d grows, in the hybrid model P targets less in order to soften competition, whereby leading S to increase both quality and price.

We can now move to compare P's profits across business models. Two intuitive effects determine P's choice. First, under the hybrid model, P introduces its private label and, therefore, (other things being equal) it has an additional source of profit compared to the pure intermediation model in which it only extracts a share of S's profit. Second, by introducing its private label, P also creates competition in the market, which lowers S's profit and, hence, the surplus that P can extract from S. When d is small, the first effect dominates. For large d, instead, the effect is ambiguous and depends on b, which de facto measures the weight attributed by P to the loss of profit associated with the intensified competition in the marketplace.

Comparing (4) with (9) it is immediate to see that, for given precision η, the following is true

$$\pi^{\star}(\eta) \geq \pi^{*}(\eta) \quad \Leftrightarrow \quad \Gamma(b, d) \geq \frac{b}{6}.$$

Then, the following can be stated:

Proposition 7 *Holding η constant across the two business models, P prefers to operate as a pure intermediary when consumers perceive products as relatively close substitutes (d large), and when its bargaining power is relatively strong (b large).*

The intuition is as follows. For given information accuracy, introducing a private label in the marketplace is not worthwhile for the platform when P's and S's products are close substitutes because competition erodes both the profit earned through its private label and S's profit. This effect becomes even stronger, in relative terms, when b grows large because in the pure intermediation model, P extracts a relatively higher share of S's monopoly profit.

What happens if the optimal accuracy varies across the two business models? There are two additional effects at play over and above those discussed before. First, other things being equal, the business model that induces information gathering features better targeted pricing and quality, thereby benefitting P. Second, since consumers have privacy concerns, demand falls in the regime in which P is informed, which ceteris paribus lowers its profit.

Bearing these effects in mind, we now investigate each of the two possible scenarios—i.e., $\eta^\star = 0 < \eta^* = 1$ and $\eta^\star = 1 > \eta^* = 0$.

Scenario 1 $\eta^\star = 0 < \eta^* = 1$. In this case, the relevant difference is

$$\pi^\star(0) - \pi^*(1) \triangleq 3A^2\Gamma(b,d) - \frac{b\big(3\psi^2 + \sigma^2 + 3A(A - 2\psi)\big)}{6} + c.$$

This expression is decreasing in σ^2 and increasing in ψ. Intuitively, the higher volatility of demand, the less appealing the hybrid model is when it induces P to be uninformed because it cannot target consumers and price discriminate them. By contrast, the hybrid model becomes more suitable when consumers feature greater privacy concerns. Moreover, since the function $\Gamma(b,d)$ is increasing in b (for given d) and decreasing in d (for given b), the impact of d is negative, meaning that a higher product substitutability tends to reduce the incentive of the platform to adopt a hybrid model: the more competition intra-platform, the lower the incentive to create rivalry with the private label. The effect of b is ambiguous: for given level of product differentiation, the higher the platform's bargaining power vis-à-vis the seller, the higher its incentive to adopt a hybrid model because it can extract a higher share of the seller's profit. Yet, the higher b, the greater the incentive to monopolize the market through the choice of a pure intermediation model. This second effect, becomes more prominent when demand is more uncertain (i.e., for higher values of σ^2) and less prominent when consumers' privacy concerns are stronger (i.e., for higher values of ψ).

Hence, when $b \to 0$, the platform prefers the hybrid model since the pure-intermediation model guarantees relatively small profits (recall that $\Gamma(0,d) > 0$). Consider thus the polar case in which $b \to 1$. In this case, we know that $c^\star < c^*$ for $\psi \leq \psi^*$ and $d \geq d^*$. As a result, $\eta^\star = 0$ and $\eta^* = 1$ for any $c \in [c^\star, c^*]$. Figure 3 plots the difference $\pi^\star(0) - \pi^*(1)$ at $c = 0$ (which is a conservative scenario since we know that $c > 0$ in the region of parameters under consideration) in the space

$(A, d) \in [2, +\infty) \times [d^*, 0.5)$ for $\psi = \sigma = 1$ (recall that $A \geq \sigma + \psi$, hence we need $A \geq 2$).

The figure shows that $\pi^\star(0) > \pi^*(1)$ in the relevant region of parameters. Hence, for large values of product substitutability, it is well possible that the preferred business model for the platform is the hybrid one and that this model minimizes the incentive to collect and disclose information, whereby protecting privacy. Yet, since $\pi^\star(0) - \pi^*(1)$ is decreasing in σ^2 the incentive to choose a hybrid platform and be uninformed drops when the market becomes relatively more volatile—i.e., when σ^2 rises.[14]

Scenario 2 $\eta^\star = 1 > \eta^* = 0$. In this case, the relevant difference is $\pi^*(0) - \pi^\star(1)$ —i.e., this comparison applies when $d < d^*$ and for any $c \in [c^\star, c^*]$. It can be checked that

$$\pi^*(0) - \pi^\star(1) \triangleq b\frac{A^2}{2} - \left(3\psi^2 + \sigma^2 + 3A(A - 2\psi)\right)\Gamma(b, d) + c,$$

which, as discussed above, is increasing in d and decreasing in ψ and σ^2. The impact of b is again ambiguous for the same reasons previously illustrated. For $b \to 0$, the expression above is clearly negative. Therefore, let us consider again the polar case $b = 1$ and $d < d^*$ so that $c^* > c^\star$. Setting $c = 0$ as before, in Fig. 4 we plot the above difference in the space $(A, d) \in [2, +\infty) \times [0, d^*)$ for $\sigma = \psi = 1$.

The figure shows that the difference $\pi^*(0) - \pi^\star(1)$ is always positive for low values of A irrespective of d and for high values of d when A grows large. The reason why the difference is positive for b large is simple: intense intra-platform competition dissipates the sellers' profits and therefore makes the pure intermediation model more appealing than the hybrid model. The reason why a higher A makes the hybrid model more appealing for intermediate values of d is due to the fact that the representative consumer features preference for variety, hence supplying more than one product in the platform is profitable when the consumer's willingness to pay is sufficiently high.

3.5 Consumer surplus

What is the effect of the business model on consumer surplus? There are three intuitive effects that shape this relationship. First, recall that the hybrid model is positively biased towards consumer surplus. In fact, under this business model, there is one additional variety compared to the pure intermediation model, and consumers like to have more options (because the Sing and Vives (1984), utility function exhibits preference for variety). Second, irrespective of P's information policy, the hybrid model yields a more competitive outcome that goes in the direction of increasing consumer surplus. Finally, privacy concerns also matter for consumers—i.e., other things being equal, they prefer the model that minimizes the incentive of the intermediary to collect information.

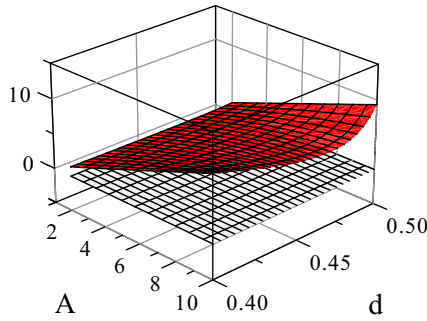[14] The same result obtains for different parameter configurations (see the Appendix).

Fig. 3 Information acquisition and profits

Building on these insights, in this section, we study the effect of the business model on consumer surplus. To neutralize the effect of product variety, we focus only on product 1 and study the effect of the business model on the part of the (representative) consumer's expected utility function that pertains to this product, and not P's product. Of course, it should be noted that such a restriction will somewhat bias results in favor of the pure intermediation model. Specifically, we compare (expected) consumer surplus in the pure intermediation model,

$$U(\eta) \triangleq \int_{-\sigma}^{\sigma} \mathbb{E}_s \left[\left(A + \theta + x_1^*(s)\right)q_1^*(s) - \frac{1}{2}q_1^*(s)^2 - p_1^*(s)q_1^*(s) - \eta\psi q_1^*(s)|\theta \right] \frac{d\theta}{2\sigma}.$$

(10)

with the following expression,

$$V(\eta) \triangleq \int_{-\sigma}^{\sigma} \mathbb{E}_s \left[\left(A + \theta + x_1^\star(s)\right)q_1^\star(s) - \frac{1}{2}q_1^\star(s)^2 - p_1^\star(s)q_1^\star(s) - \eta\psi q_1^\star(s)|\theta \right] \frac{d\theta}{2\sigma},$$

(11)

which only accounts for the utility that consumers derive from product 1 in the hybrid model as it happens in the pure intermediation model.

To gain insights on how P's information policy affects consumer surplus, a first useful exercise is to plot the above expressions for different levels of the cost of privacy ψ. Figure 5 shows that they are decreasing in η for ψ large and increasing or inverted-U shaped for low values of ψ.[15]

Under both business models, an information policy with higher accuracy harms consumers when privacy concerns are strong enough (ψ large), whereas a positive level of accuracy benefits consumers when privacy concerns are not too important (ψ sufficiently low). The reason why consumers may prefer to disclose some information about their willingness to pay is clear: this information allows sellers to target price and quality more accurately—i.e., it avoids pricing high and setting quality high when demand is high and vice-versa.

---

[15] Similar patterns arise for alternative parameter configurations (as reported in the Appendix).
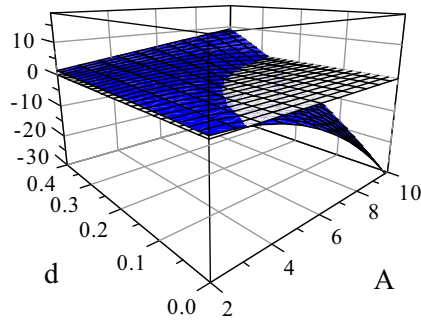
Fig. 4 Information acquisition and profits

With the above effects in mind, we can now compare expected utilities in the two business model. As before, it is useful to start the analysis by holding the level of accuracy constant across the two business models. Figure 6 simulates the difference $V(\eta) - U(\eta)$ as a function of d for alternative parametric restrictions.

Hence, even when neutralizing the effect of the introduction of an additional variety in the hybrid model, conditional on the platform choosing the same information policy in both business models, consumer surplus is higher in the hybrid model than in the pure intermediation model. The reason is simple: for a given accuracy, in the hybrid model, S may loose business to P, which leads to a more competitive outcome.

Figure 7, instead, plots the difference between consumer surplus in the two models for different levels of accuracy.

Panel A shows that the hybrid model still performs better for consumers when P does not collect information in the hybrid mode ($\eta^\star = 0$) and it does so in the pure intermediation model ($\eta^* = 1$). Panel B illustrates the opposite scenario in which P collects information in the hybrid model and does not in the pure intermediation model: in this case, the hybrid model under-performs for low levels of d (weak competition or relatively differentiated products), while it still benefits consumers for relatively high values of d (intense competition or relatively close substitutes).

4 Extensions, robustness and further remarks

We now discuss some extensions of the baseline analysis to check its robustness and further issues that have been not addressed in the above analysis.

4.1 Alternative demand function

Instead of assuming the modified version of the Singh and Vives (1984) demand structure used in the baseline model, we consider now the following alternative specification for the representative consumer's utility function
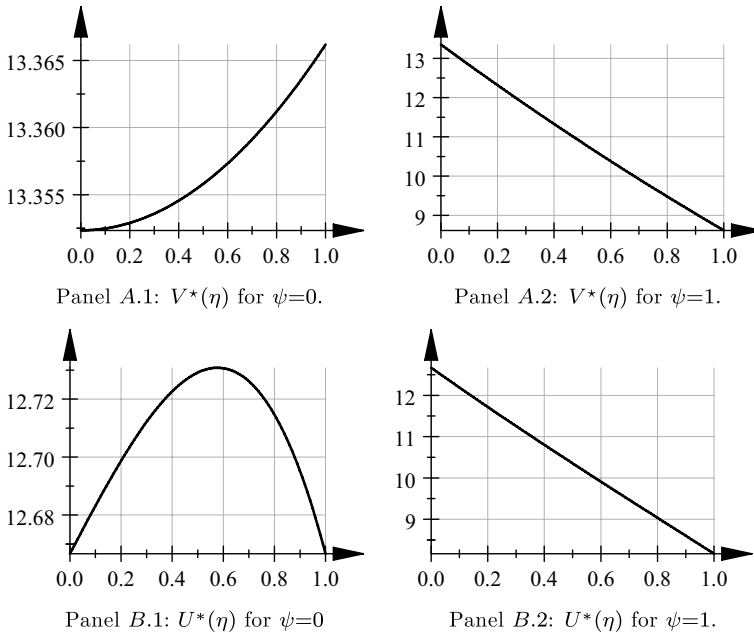
Panel $A.1$: $V^\star(\eta)$ for $\psi=0$.

Panel $A.2$: $V^\star(\eta)$ for $\psi=1$.

Panel $B.1$: $U^*(\eta)$ for $\psi=0$

Panel $B.2$: $U^*(\eta)$ for $\psi=1$.

Fig. 5 Simulations of CS for given business model

$$U(\cdot) \triangleq \underbrace{\sum_{i=1}^{N} \left(A + \theta + x_i\right)q_i - \frac{d}{2(1+d)}\left(\sum_{i=1}^{N} q_i\right)^2 - \frac{N}{2(1+d)} \sum_{i=1}^{N} q_i^2 - \sum_{i=1}^{N} p_i q_i}_{\text{Shubick-Levitan Utility}} - \underbrace{\eta\psi \sum_{i=1}^{N} q_i}_{\text{Privacy Concerns}},$$

where, as before, $N = 1$ in the pure intermediation model, and $N = 2$ in the hybrid model.

While privacy concerns are still as in the baseline model, the first part of the above expression corresponds to the standard Shubik-Levitan quadratic utility function (see, e.g., Motta, 2004, Ch. 8.4.2.), which is often used in the IO literature as an alternative specification to Singh and Vives (1984). The parameter $d \geq 0$ captures again the degree of product differentiation between products. As in the baseline model, we assume d not too large to avoid corner solutions—i.e., $d \leq \frac{5}{2}$.

A somewhat nice property of using this alternative specification is that the analysis of the pure intermediation model does not change since both utility functions yield the same demand when $N = 1$. Hence, we focus on the hybrid model. Differentiating the above utility function with respect to quantities and solving the corresponding system of first-order conditions we have the following demand functions

$$q_i\left(p_i, p_j, x_i, x_j, \theta\right) \triangleq \frac{A + \theta - \psi\eta + x_i - p_i}{2} + d\frac{(p_j - p_i) - (x_j - x_i)}{4} \quad \forall i,j = 1,2.$$
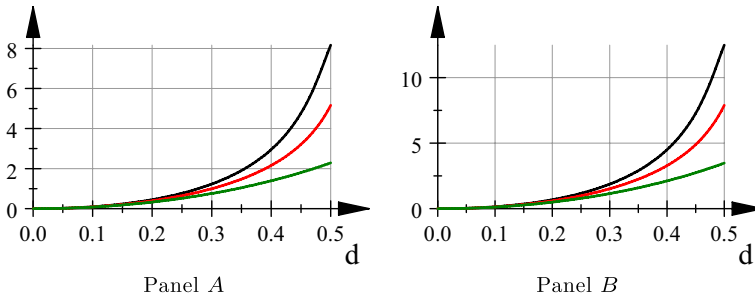
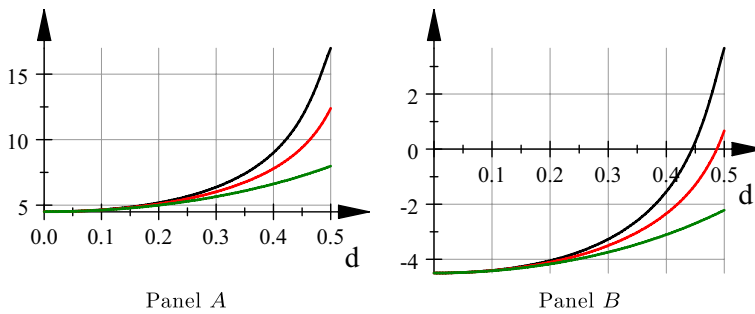**Fig. 6** Difference in CS for given information acquisition decision

Fig. 7 Difference in CS for different information acquisition decisions

Following again a backward induction approach in solving the game (see the Appendix) we can state the following result.

Proposition 8 *There exists a positive function $\hat{\Gamma}(b,d) \geq 0$ such that*

$$\hat{\pi}^{\star}(s) \triangleq [A + \eta(s - \psi)]^2 \Gamma_{SL}(b,d).$$

Hence, P's expected profit net of the cost of collecting information with accuracy η is

$$\hat{\pi}^{\star}(\eta) \triangleq \left(\eta^2\left(3\psi^2 + \sigma^2\right) + 3A(A - 2\psi\eta)\right)\Gamma_{SL}(b,d) - c\eta. \qquad (12)$$

The optimal accuracy is such that $\eta^{\star} = 1$ only if $\psi \leq \psi^*$ and if

$$c \leq c_{SL}^{\star} \triangleq \left(\sigma^2 - 3\psi(2A - \psi)\right)\Gamma_{SL}(b,d).$$

Otherwise, $\eta^{\star} = 0$.

The intuition is the same as in the baseline model. Therefore, comparing the outcome above with the results obtained for the pure intermediation model, we have

$$c_{SL}^{\star} \gtrless c^{*} \quad \Leftrightarrow \quad \Gamma_{SL}(b,d) \gtrless \frac{b}{6}.$$

We can thus state the following.

Proposition 9 *When $\psi \geq \psi^{*}$ the signal is uninformative irrespective of the business model. By contrast, when $\psi < \psi^{*}$, there exists a threshold $d_0 \in \left(0, \frac{5}{2}\right)$ such that*:

- if $d \leq d_0$ then $\eta^{\star} = 1$ and $\eta^{*} = 0$ for every $b \in [0,1]$.
- if $d > d_0$ then there exists a function $b_0(d) < 1$, with $b_0'(d) < 0$, such that $\eta^{\star} = 0$ and $\eta^{*} = 1$ if and only if $b \geq b_0(d)$.

This result confirms that the findings of the baseline model extend to different consumer preferences. Figure 8 illustrates the result graphically by plotting the difference $\Gamma_{SL}(b,d) - \frac{b}{6}$ (red surface) in the space $(b,d) \in [0,1] \times [0, \frac{5}{2})$.

The figure shows that the difference $c_{SL}^{\star} - c^{*}$ is negative when b and d are both sufficiently large. Hence, P has a weaker incentive to collect information under the hybrid model than in the pure intermediation model when intra-platform competition is relatively intense (i.e., when d is sufficiently large) and when P has strong bargaining position vis-à-vis S (i.e., when b is sufficiently large).

4.2 Decentralized market decisions

So far, we assumed that P and its retail unit (hereafter R) are vertically integrated and that the pricing and the marketing decisions are centralized at the platform level—i.e., they are chosen to maximize P 's and R's joint profit. We now assume, on the contrary, that R has full discretion in choosing product 2's price and quality. As in the baseline model, we posit again the Singh and Vives (1984) utility function.

While the pure intermediation model is unaffected by this hypothesis, in the hybrid model the equilibrium is symmetric because R only cares about product 2's profit. The optimal accuracy (see the Appendix) is 1 only if $\psi \leq \psi^{*}$ and if

$$c \leq c_D^{\star} \triangleq \left(\sigma^2 - 3\psi(2A - \psi)\right)\Gamma_D(b,d),$$

with

$$\Gamma_D(b,d) \triangleq \frac{(1+b)(1-2d^2)}{6(1+d(1-d))^2}.$$

In Fig. 9 we compare the incentives to collect information in the two business models by plotting the difference $\Gamma_D(b,d) - \frac{b}{6}$.

Hence, the qualitative results of the baseline model hold even with vertical delegation. Interestingly, by comparing Figs. 2 and 9 it can be seen that the incentive of the hybrid platform to remain uninformed is stronger when it delegates the pricing and the targeting decisions to the retail unit. In other words, platforms where
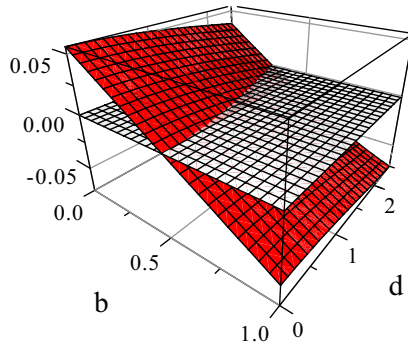
Fig. 8 Incentives to acquire information

decision rights are likely to be delegated to their retail units, have weaker incentives to gather and disclose information than platforms featuring a highly hierarchal organization and centralized decision 'nodes'.

4.3 Multiple sellers

The case of multiple sellers competing in the marketplace is rather intuitive. Consider for example a representative consumer with preferences defined by the following utility function

$$U(\cdot) \triangleq \sum_{i=1}^{M} (A + \theta + x_i)q_i - \frac{1}{2}\sum_{i=1}^{M} q_i^2 - d\sum_{j=1}^{M}\sum_{i\neq j}^{M} q_i q_j - \sum_{i=1}^{M} p_i q_i - \eta\psi M\sum_{i=1}^{M} q_i,$$
(13)

where $M = N$ in the pure intermediation model, and $M = N + 1$ in the hybrid model.[16]

Notably, dealing with more than one seller allows us to introduce an aspect that we had neglected so far. Precisely, we now assume that privacy concerns are increasing with the number N of sellers (each denoted by S_i, with $i = 1, .., N$) to which P discloses information (again, we assume *no discrimination* in the sense that the information collected by P must be disclosed to all sellers listed in the platform).[17] The idea is that the more sellers manage consumer personal data, the more likely it is that some information lands in the wrong hands, whereby hurting consumers.

In this framework, we show (see the Appendix) that when b, d and N are sufficiently large, P has no incentive to gather and disclose information in the hybrid model, whereas it gathers information when acting as a pure intermediary. The intuition is essentially the same as in the baseline model. Indeed, an increased number of varieties is just another way of measuring competition within the marketplace: the

---

[16] As before, we assume that P's private label is identified with product $N + 1$, and its quantity is q_{N+1}.

[17] Notice that we did not introduce this assumption in the baseline model in order no to bias the results in favour of the hybrid business model.
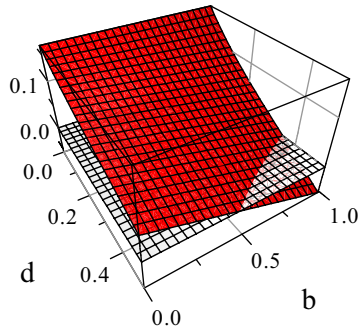
Fig. 9 Impact of b and d on the incentive to acquire information

more varieties there are, the lower will be the market share of each seller, and therefore the more intense competition will be.

One interesting aspect has to do with the impact of the number of varieties on the incentive to gather information within each business model. Two main contrasting forces are at play. First, when the number of varieties increases, other things being equal, the platform is more willing to gather information and allow sellers to make more accurate decisions, so to extract a higher profit from them. Second, recall that in this framework, privacy concerns are increasing in N. Hence, when the number of sellers is too large, gathering information triggers significant privacy concerns, whereby refraining the platform from being informed. When N is small, the first effect dominates; otherwise, the second effect dominates. Hence, while for large values of d the effect of an increase in N is likely to be negative in both business models, for small values of d there is an inverted-U shaped pattern (see the Appendix).

4.4 Endogenous listing fees

So far, we have considered an efficient profit-sharing (bargaining) rule to isolate the effects of information gathering from those of double marginalization. Yet, in reality, platforms charge either per-unit fees (i.e., sellers pay platforms a fee for each unit distributed through the platform) or ad-valorem fees (i.e., they extract a percentage of the price charged to final consumers). In both cases, a double marginalization phenomenon occurs at equilibrium. The extent to which information gathering affects the seller's incentive to pass on the listing fee to final consumers determines another channel through which the platform's business model alters the incentive to gather information. In what follows, we discuss how the introduction of these contractual frictions is likely to change our results.

Ad-valorem fee. In the online Appenix we show that with an ad-valorem feee P has a greater incentive to gather information in the hybrid model than in the pure intermediation model, and this incentive drops as products become closer substitutes. The reason is as follows. As standard in the literature, with zero marginal cost

downstrea, *P* fully extracts *S*'s revenue with an ad-valorem fee, so that *S* does not invest in quality/advertising. This fact grants *P* a competitive advantage in the sense that the demand for its private label is not reduced by the rivals' investment, and thus the information that it collects will be used only to target quality/advertising of its own brand without fearing that the rival will do the same. Of course, when products become closer substitutes, competition intensifies in the hybrid marketplace and *S* uses the information received from *P* to set prices more competitively, thereby mitigating *P*'s incentive to gather information.

What would happen with a positive marginal cost? In this case, the optimal ad-valorem fee will typically be lower than 1.[18] Therefore, *S* will earn a positive profit margin and invest (i.e., $x_2 > 0$). This will, in turn, spoil *P*'s competitive advantage discussed above and make results closer to the baseline model. That is, *P* may have a stronger incentive to acquire information in the pure intermediation model than in the hybrid model when *d* is sufficiently large: in this case, *S* will compete more aggressively on both dimensions (marketing and pricing), which may reverse the above result.

Linear per-unit fee The same logic illustrated above applies with a linear per-unit fee. The hybrid model yields a higher incentive to gather information. Interestingly, in contrast to the case of an ad-valorem fee, this difference is U-shaped with respect to *d* (see the online Appendix). As the degree of product differentiation varies, there are two effects that shape *P*'s relative incentive to gather information. When products are sufficiently differentiated (*d* small), an increase in *d* makes *P* relatively less willing to gather information under the hybrid model because *S* will use this information to set price and quality more competitively. When instead products are close substitutes (*d* large), prices are not too distant one from the other, and *P* becomes relatively more willing to collect information under the hybrid model because this information will be used by *S* to target demand and create more vertical differentiation, which ceteris paribus benefits *P* too.

5 Conclusions

Consumer privacy is at the heart of the policy debate on the benefits and costs of the digital revolution. The extensive use and exchange of consumer data call for a deeper understanding of the forces shaping platforms' incentives to collect such information.

In this paper, we have contributed to the growing literature on platforms' organization by investigating the role of business models as drivers of the accuracy of consumer information collected by platforms. We have argued that there is no objective presumption that pure intermediation platforms collect more or less information than hybrid platforms. Platforms' business model is not neutral to data collection. This observation should be considered carefully by privacy authorities, especially because in the EU the GDPR is based on the 'data minimization principle.' Specifically, we have concluded that the relationship between the business model and the

---

[18] This is immediate since *S* would make negative profit margins at $\tau = 1$.

platform's incentive to collect consumer information depends on the degree of intra-platform competition and platforms' bargaining strength vis-à-vis third-party sellers. Pure intermediation platforms are less likely to collect and disclose information than hybrid ones when their bargaining position is not too strong—e.g., because of intense inter-platform competition—and their private labels are highly differentiated from their third-party rivals' products; the opposite holds otherwise.

We have made a few assumptions to obtain these results and have thus purposefully neglected some relevant aspects of platform markets. For example, we have neglected platform competition and ecosystem effects—e.g., investments made by platforms to benefit all participants to their marketplaces but at the expense of competing platforms. These two aspects are gaining growing importance in online markets, especially regarding the difference between open and closed ecosystems/platforms. We have also assumed away repeated interaction and collusion, a prominent issue in digital markets in light of the recent developments on algorithmic collusion (e.g., Calvano et al., 2020). Finally, while we have primarily focused on an efficient contracting rule to identify the pure relationship between information acquisition and the platform's business models, our results suggest that the presence of double marginalization (as induced by inefficient contracting rules) may impact such a relationship in a more complex way. All these aspects are on our research agent, and we hope to explore them and other additional interesting elements of platform markets soon.

Appendix

Proof of Proposition 1 Taking the difference between $\pi^*(1)$ and $\pi^*(0)$, we have

$$\pi^*(1) - \pi^*(0) = \frac{b(\sigma^2 - 3\psi(2A - \psi))}{6} - c,$$

which in the relevant range of parameters is negative if and only if

$$\psi \geq \psi^* \triangleq A - \sqrt{A^2 - \frac{\sigma^2}{3}}.$$

Hence, $\pi^*(1) \geq \pi^*(0)$ only if $\psi \leq \psi^*$ and if

$$c \leq c^* \triangleq \frac{b(\sigma^2 - 3\psi(2A - \psi))}{6},$$

which shows the result. □

Proof of Proposition 2 With an hybrid platform, the market equilibrium is obtained by solving the system of first-order conditions (5)–(8). Moreover, it can be shown that

$$p_2^\star(s) - p_1^\star(s) = -\frac{bd^2(1-d^2)(A + \eta(s - \psi))}{1 - bd^2(1-d^2) - d^2(3 - d^2)} \leq 0,$$

and that

$$x_2^\star(s) - x_1^\star(s) = \frac{1}{1 - d^2}\left(p_1^\star(s) - (1 + bd)p_2^\star(s)\right),$$

which is negative since $p_1^\star(s) \leq p_2^\star(s)$. $\qquad\square$

Proof of Lemma 3 Conditional on observing signal s, P's expected profit is

$$\pi^\star(s) \triangleq \mathbb{E}\left[q_2\left(p_2^\star(s), p_2^\star(s), x_2^\star(s), x_1^\star(s), \theta\right)|s\right]p_2^\star(s) - \frac{x_2^\star(s)^2}{2}$$
$$+ b\left[\mathbb{E}\left[q_1\left(p_1^\star(s), p_2^\star(s), x_1^\star(s), x_2^\star(s), \theta\right)|s\right]p_1^\star(s) - \frac{x_1^\star(s)^2}{2}\right].$$

It can be shown that

$$\pi^\star(s) = (A + \eta(s - \psi))^2\Gamma(b, d),$$

with

$$\Gamma(d, b) \triangleq \frac{(1 - 2d^2)(1 - d(1 + d))^2}{6(1 - bd^2(1-d^2) - d^2(3 - d^2))^2}$$
$$+ \frac{(1 - d(1 + d))(1 - d - 5d^2 + 4d^3 + 4d^4 - 2d^5)b}{6(1 - bd^2(1-d^2) - d^2(3 - d^2))^2}$$
$$- \frac{d^2(1 - 2d)(1 - d)^2(1 + d)^2b^2}{6(1 - bd^2(1-d^2) - d^2(3 - d^2))^2}.$$

Hence, integrating $\pi^\star(s)$ with respect to s we have

$$\pi^\star(\eta) \triangleq \int_{-\sigma}^{\sigma} \pi^\star(s)\frac{ds}{2\sigma} - c\eta = \left(\eta^2(3\psi^2 + \sigma^2) + 3A(A - 2\psi\eta)\right)\Gamma(b, d) - c\eta,$$

which establishes the result. $\qquad\square$

Proof of Proposition 4 Comparing $\pi^\star(1)$ with $\pi^\star(0)$ we have

$$\pi^\star(1) - \pi^\star(0) = \left(\sigma^2 - 3\psi(2A - \psi)\right)\Gamma(b, d) - c,$$

Hence, $\pi^\star(1) \geq \pi^\star(0)$ only if $\psi \leq \psi^*$ and if

$$c \leq c^\star \triangleq \left(\sigma^2 - 3\psi(2A - \psi)\right)\Gamma(b, d),$$

which concludes the proof. $\qquad\square$

Proof of Proposition 6 Showing that if $\psi \geq \psi^*$ then P does not collect information regardless of the business model is straightforward. Consider then $\psi < \psi^*$. Using the expression for $\Gamma(d, b)$ we have

$$\Gamma(d, 0) = \frac{1 - 2d^2}{6(1 + d(1 - d))^2} > 0,$$

which immediately implies that for $b \to 0$ the incentive to collect information is stronger in the hybrid model than in the pure intermediation model for every admissible level of d.

Next, using the expression for $\Gamma(d, b)$ again

$$\Gamma(d, 1) = \frac{2 - 4d - 9d^2 + 18d^3 + 10d^4 - 18d^5 - 5d^6 + 4d^7}{6(1 - 4d^2 + 2d^4)^2},$$

so that, for $\psi \leq \psi^*$, it follows that

$$c^\star \geq c^* \quad \Leftrightarrow \quad \Gamma(d, 1) \geq \frac{1}{6},$$

which is true for $d \geq d^* \approx 0.4$ (as shown in Fig. 2). $\qquad\square$

Proof of Corollary 6 The proof follows immediately by inspection of Fig. 2.

Consumer surplus The simulation done for consumer surplus obtains by substituting the equilibrium values of prices and advertising levels into Eqs. (11) and (10). We find that

$$U(\eta) = \frac{\eta + 3d^2\eta^2 - 30d\eta - \eta^3 + 76}{6},$$

which is plotted for $\psi = 0$ (Panel B.1) and $\psi = 1$ (Panel B.2) in Fig. 4.

By the same token we find that, in the same region of parameters,

$$V(\eta) = 13.352 + 0.52763d^2\eta^2 - 5.2763d\eta + \frac{1.3839}{100}\eta^2,$$

which is plotted for $\psi = 0$ (Panel A.1) and $\psi = 1$ (Panel A.2) in Fig. 4.

For alternative values of b, in Fig. 5 we plot the differences

$$
\begin{aligned}
V(0) - U(0) = {} & \frac{d^2 (74 + 298d + 301d^2 + 2d^3 - 76d^4)(1 - d(1 + d))^2}{6(1 + d)^2 (1 - bd^2 (1 - d^2) - d^2 (3 - d^2))^2} \\
& + \frac{2d^3(1 - d)(d + 1)(1 - d(1 + d))(75 + 76d - 74d^2 - 76d^3)b}{6(1 + d)^2 (1 - bd^2 (1 - d^2) - d^2 (3 - d^2))^2} \\
& - \frac{d^4 (150d + 76d^2 + 75)(1 - d)^2(1 + d)^2b^2}{6(1 + d)^2 (1 - bd^2 (1 - d^2) - d^2 (3 - d^2))^2},
\end{aligned}
$$

and

$$V(1) - U(1) = \frac{49d^2\left(1 + 2d - d^2\right)(1 - d(1 + d))^2}{6\left(1 - bd^2\left(1 - d^2\right) - d^2\left(3 - d^2\right)\right)^2}$$
$$+ \frac{49d^2 2bd(1 + d)(1 - d(1 + d))(1 - d)^2}{6\left(1 - bd^2\left(1 - d^2\right) - d^2\left(3 - d^2\right)\right)^2}$$
$$- \frac{49b^2 d^4 (1 - d)^2 (1 + d)^2}{6\left(1 - bd^2\left(1 - d^2\right) - d^2\left(3 - d^2\right)\right)^2}.$$

Finally, for alternative values of b, in Fig. 6 we plot the differences

$$V(0) - U(1) = \frac{\left(27 + 108d + 182d^2 + 244d^3 + 193d^4 + 2d^5 - 49d^6\right)(1 - d(1 + d))^2}{6\left(1 - bd^2\left(1 - d^2\right) - d^2\left(3 - d^2\right)\right)^2}$$
$$- \frac{2bd^2(1 - d)(d + 1)(1 - d(1 + d))\left(6d - 22d^2 + 47d^3 + 49d^4 + 27\right).}{6\left(1 - bd^2\left(1 - d^2\right) - d^2\left(3 - d^2\right)\right)^2}$$
$$- \frac{b^2 d^4 \left(96d + 49d^2 + 48\right)(1 - d)^2(1 + d)^2.}{6\left(1 - bd^2\left(1 - d^2\right) - d^2\left(3 - d^2\right)\right)^2}$$

and

$$V(1) - U(0) = -\frac{\left(27 + 54d - 76d^2 - 152d^3 + 76d^4\right)(1 - d(1 + d))^2}{6\left(1 - bd^2\left(1 - d^2\right) - d^2\left(3 - d^2\right)\right)^2}$$
$$- \frac{2bd^2(1 - d)(1 + d)\left(-76d + 76d^2 - 27\right)(1 - d(1 + d))}{6\left(1 - bd^2\left(1 - d^2\right) - d^2\left(3 - d^2\right)\right)^2}$$
$$- \frac{76b^2 d^4 (1 - d)^2 (d + 1)^2}{6\left(1 - bd^2\left(1 - d^2\right) - d^2\left(3 - d^2\right)\right)^2}.$$

Proof of Proposition 9 Solving the system of first-order conditions we have

$$p_1^\star(s) \triangleq \frac{4(2(3 + d) + d(1 - bd) + d(b - d))}{36 + 20d + 4d(1 - bd) - d^2(3 + d)}(A + \eta(s - \psi))$$
$$\leq p_2^\star(s) \triangleq \frac{4(2(3 + d) + d(1 - d))}{36 + 20d + 4d(1 - bd) - d^2(3 + d)}(A + \eta(s - \psi)),$$

$$x_1^\star(s) = \frac{12(1 + d) - 4bd(1 + d) + d^2(1 - d)}{36 + 20d + 4d(1 - bd) - d^2(3 + d)}(A + \eta(s - \psi))$$
$$\leq x_2^\star(s) = \frac{(2 + d)(2(3 + d) + d(1 - d))}{36 + 20d + 4d(1 - bd) - d^2(3 + d)}(A + \eta(s - \psi)).$$

Substituting these values into P's expected profit, we have

$$\pi_{SL}^{\star}(\eta) \triangleq \left(\eta^2\left(3\psi^2 + \sigma^2\right) + 3A(A - 2\psi\eta)\right)\Gamma_{SL}(b, d) - c\eta$$

with

$$\Gamma_{SL}(d, b) \triangleq \frac{(d + 2)(6 - d)(6 + d(3 - d))^2}{6\left(36 + 20d + 4d(1 - bd) - d^2(3 + d)\right)^2}$$
$$+ \frac{(6 + d(3 - d))\left(2(36 + 19d) + 22d(1 - d) - d^3(7 - d)\right)b}{6\left(36 + 20d + 4d(1 - bd) - d^2(3 + d)\right)^2}$$
$$- \frac{6d^2(1 + d)(3 - d)b^2}{6\left(36 + 20d + 4d(1 - bd) - d^2(3 + d)\right)^2}.$$

The function $\pi_{SL}^{\star}(\eta)$ is again convex in η. Hence, comparing $\pi_{SL}^{\star}(1)$ with $\pi_{SL}^{\star}(0)$ we obtain that $\eta_h^{\star} = 1$ only if

$$\psi \leq \psi^* \triangleq A - \sqrt{A^2 - \frac{\sigma^2}{3}},$$

and if

$$c \leq c_{SL}^{\star} \triangleq \left(\sigma^2 - 3\psi(2A - \psi)\right)\Gamma_{SL}(d, b).$$

The proof of the result is then directly implied by Fig. 8. $\qquad\square$

Decentralized decisions and the value of information First, notice that in the pure intermediation model nothing changes. By contrast, in the hybrid model, with decentralized decisions, the equilibrium of the market game is symmetric—i.e.,

$$p^{\star}(s) = \frac{\left(1 - d^2\right)(1 - d(1 + d))(A + \eta(s - \psi))}{1 - d^2\left(3 - d^2\right)},$$

and

$$x^{\star}(s) = \frac{(1 - d(1 + d))(A + \eta(s - \psi))}{1 - d^2\left(3 - d^2\right)}.$$

Hence,

$$\pi_D^{\star}(s) \triangleq \mathbb{E}\left[q_1(\cdot) + bq_2(\cdot)|s\right]p^{\star}(s) - (1 + b)\frac{x^{\star}(s)^2}{2} = [A + \eta(s - \psi)]^2\Gamma_D(b, d),$$

with

$$\Gamma_D(d,b) = \frac{(1+b)(1-2d^2)}{6(1+d(1-d))^2}.$$

Substituting $p^\star(s)$ and $x^\star(s)$ into $\pi_D^\star(s)$ and integrating, we have

$$\pi_D^\star(\eta) = \int_{-\sigma}^{\sigma} \pi_D^\star(s)\frac{ds}{2\sigma} - c\eta = \left(\eta^2\left(3\psi^2 + \sigma^2\right) + 3A(A - 2\psi\eta)\right)\Gamma_D(d,b) - c\eta.$$

Hence, when P delegates the pricing and advertising decisions to R, it collects and discloses information only if $\psi \leq \psi^*$ and if

$$c \leq c_D^\star \triangleq \left(\sigma^2 - 3\psi(2A - \psi)\right)\Gamma_D(d,b).$$

In Fig. 8 we plot the difference between $\Gamma_D(d,b)$ and $\frac{b}{6}$, which shows immediately that the results of the baseline model hold qualitatively also with decentralized decision making.

Multiple sellers Differentiating the utility function (13) with respect to quantities, we obtain the following system of first-order conditions

$$A + \theta + x_i - q_i - d\sum_{j\neq} q_j - p_i - \eta M\psi = 0 \quad \forall i = 1,..,M,$$

whose solution yields the system of direct demand functions

$$q_i(\cdot) = \frac{A + \theta - M\psi\eta}{1 + (M-1)d} - \frac{(1 + (M-2)d)(p_i - x_i) - d\sum_{j\neq i}(p_j - x_j)}{(1 + (M-1)d)(1-d)}, \quad \forall i = 1,..,M.$$

We assume $A - \sigma - M\psi \geq 0$ to guarantee positive prices and quantities. Recall that $M = N$ in the pure intermediation model, and $M = N + 1$ in the hybrid model. For simplicity, assume $b = 1$.

Consider first a pure intermediary. Focus on a symmetric equilibrium in which every seller chooses $p^N(s)$ and $x^N(s)$. Seller i solves the following maximization problem

$$\max_{p_i \geq 0, x_i \geq 0} \mathbb{E}\left[q_i(\cdot)|s\right]p_i - \frac{x_i^2}{2},$$

whose first-order conditions with respect to p_i and x_i, are respectively

$$\frac{A + (s - N\psi)\eta}{1 + (N-1)d} - \frac{(1 + (N-2)d)\left(2p^N(s) - x^N(s)\right) - d(N-1)\left(p^N(s) - x^N(s)\right)}{(1 + (N-1)d)(1-d)} = 0,$$

and

$$\frac{1 + (N-2)d}{(1 + (N-1)d)(1-d)}p^N(s) - x^N(s) = 0.$$

The intuition behind these conditions is the same as in the baseline model, and will be omitted for brevity. The solution is
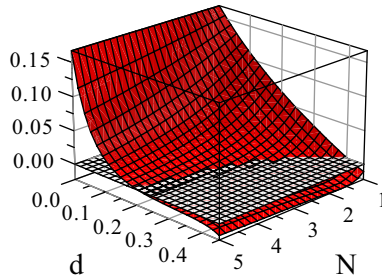
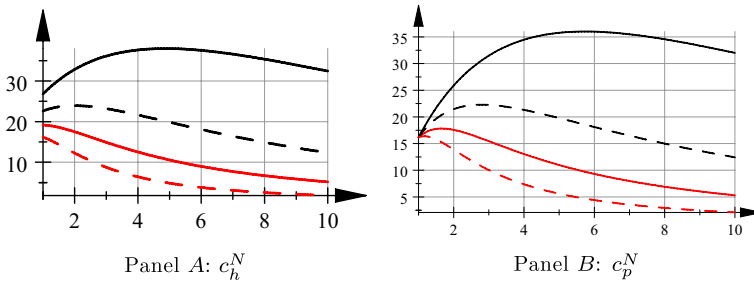**Fig. 10** Information acquisition and business model

Panel A: c_h^N

Panel B: c_p^N

Fig. 11 Simulations of critical costs

$$p^N(s) = \frac{(1-b)(1+b(N-1))(A+(s-\psi N)\eta)}{1-3b(1-b)-4b^2N+b^2N^2+2bN},$$

$$x^N(s) = \frac{(1+b(N-2))(A+(s-\psi N)\eta)}{1-3b(1-b)-4b^2N+b^2N^2+2bN}.$$

Substituting the equilibrium outcomes into P's expected profit, we have

$$N\int_{-\sigma}^{\sigma}\left\{\mathbb{E}\big[q_i(\cdot)|s\big]p_N^*(s) - \frac{x_N^*(s)^2}{2}\right\}\frac{ds}{2\sigma} - \eta c$$
$$= \left(\eta^2\left(\sigma^2+3\psi^2N^2\right)+3A(A-2A\psi N\eta)\right)\Gamma_p^N(N,d) - c\eta,$$

with

$$\Gamma_p(N,d) \triangleq \frac{\left(1-d-2Nd^2+Nd\right)(N+1)(1+(N-1)d)}{6\left(1-d-2Nd^2+N^2d^2+2Nd\right)^2}.$$

Hence, P collects information only if

$$\psi \le \psi_N^* = \frac{A}{N} - \frac{1}{N}\sqrt{A^2-\frac{\sigma^2}{3}},$$

and if

$$c \leq c_N^* \triangleq \left(\sigma^2 - 3N\psi(2A - N\psi)\right)\Gamma(N, d).$$

In the hybrid model there are $N + 1$ products distributed through the platform. Consider an equilibrium in which P sets $p_{N+1}^N(s)$ and $x_{N+1}^N(s)$, whereas its N rivals set $p_h^N(s)$ and $x_h^N(s)$.

The first-order conditions associated with each S_i's maximization problem are

$$A + (s - N\psi)\eta = \frac{(1 + (N-1)d)\left(2p_h^N(s) - x_h^N(s)\right) - d(N-1)\left(p_h^N(s) - x_h^N(s)\right) - d\left(p_{N+1}^N(s) - x_{N+1}^N(s)\right)}{1 - d},$$

and

$$\frac{1 + (N-1)d}{(1 + Nd)(1 - d)}p_h^N(s) - x_h^N(s) = 0.$$

P's maximization problem is, instead,

$$\max_{x_{N+1}, p_{N+1}} \quad q_{N+1}(\cdot)p_{N+1} - \frac{x_{N+1}^2}{2} + \sum_{i=1}^{N}\left(q_i(\cdot)p_h^N(s) - \frac{x_h^N(s)^2}{2}\right).$$

Differentiating with respect to p_{N+1} and x_{N+1}, the first-order conditions are

$$A + (s - N\psi)\eta - \frac{(1 + (N-1)d)\left(2p_{N+1}^N(s) - x_{N+1}^N(s)\right) - dN\left(p_h^N(s) - x_h^N(s)\right)}{1 - d} + \frac{bNdp_h^N(s)}{1 - d} = 0,$$

and

$$\frac{1 + (N-1)d}{(1 + Nd)(1 - d)}p_{N+1}^N(s) - x_{N+1}^N(s) - \frac{bNdp_h^N(s)}{(1 + Nd)(1 - d)} = 0.$$

The economic intuition for these conditions, which we omit for brevity, is straightforward and in line with the reasoning developed in the baseline model. The solution is

$$p_{N+1}^\star(s) = \frac{(1-d)(1+d+Nd)\left(d+3Nd^2-Nd+2d^2-1\right)(A+(s-\psi N)\eta)}{4Nd^2+6Nd^3-4Nd^4-3N^2d^2+6N^2d^3-N^3d^3+2N^3d^4-3Nd+4d^2-2d^4-1},$$

$$p_h^\star(s) = \frac{(1-b)(1+b+Nb)\left(b+2Nb^2-Nb+b^2-1\right)(A+(s-\psi N)\eta)}{4Nb^2+6Nb^3-4Nb^4-3N^2b^2+6N^2b^3-N^3b^3+2N^3b^4-3Nb+4b^2-2b^4-1},$$

$$x_{N+1}^\star(s) = \frac{\left(2d+4Nd^2-Nd^3+N^2d^3-Nd+d^2-d^3-1\right)(A+(s-\psi N)\eta)}{4Nd^2+6Nd^3-4Nd^4-3N^2d^2+6N^2d^3-N^3d^3+2N^3d^4-3Nd+4d^2-2d^4-1},$$

$$x_h^\star(s) = \frac{\left(d+2Nd^2-Nd+d^2-1\right)(Nd+1)(A+(s-\psi N)\eta)}{4Nd^2+6Nd^3-4Nd^4-3N^2d^2+6N^2d^3-N^3d^3+2N^3d^4-3Nd+4d^2-2d^4-1}.$$

Hence, substituting for the equilibrium prices and targeting levels, P's expected profit is

$$\int_{-\sigma}^{\sigma} \left\{ \mathbb{E}\big[q_{N+1}(\cdot)|s\big] p_{N+1}^N(s) - \frac{x_{N+1}^N(s)^2}{2} + \sum_{i=1}^N \left(\mathbb{E}\big[q_i(\cdot)|s\big] p_h^N(s) - \frac{x_h^N(s)^2}{2} \right) \right\} \frac{ds}{2\sigma} - c\eta$$
$$= \left(\eta^2 \big(\sigma^2 + 3\psi^2 N^2\big) + 3A(A - 2A\psi N\eta) \right) \Gamma_h(N, d) - c\eta,$$

with

$$\Gamma_h(N, d) \triangleq \frac{N - 4d - 31Nd^2 + 4N^2d + 4Nd^3 + 79Nd^4 - 34Nd^5 - 46Nd^6 + 12Nd^7 - N^2d^2 - 56N^2d^3 + 6N^3d^2 + 83N^2d^4}{6\big(4Nd^2 + 6Nd^3 - 4Nd^4 - 3N^2d^2 + 6N^2d^3 - N^3d^3 + 2N^3d^4 - 3Nd + 4d^2 - 2d^4 - 1\big)^2}$$
$$- \frac{18N^3d^3 - 50N^2d^5 + 21N^3d^4 - 4N^4d^3 + 96N^2d^6 - 98N^3d^5 + 19N^4d^4 + 49N^3d^6 - 20N^4d^5 - N^5d^4}{6\big(4Nd^2 + 6Nd^3 - 4Nd^4 - 3N^2d^2 + 6N^2d^3 - N^3d^3 + 2N^3d^4 - 3Nd + 4d^2 - 2d^4 - 1\big)^2}$$
$$- \frac{30N^3d^7 - 19N^4d^6 + 6N^5d^5 + 30N^4d^7 - 12N^5d^6 + 8N^5d^7 - 6Nd + 9d^2 - 18d^3 - 10d^4 + 18d^5 + 5d^6 - 4d^7 - 2}{6\big(4Nd^2 + 6Nd^3 - 4Nd^4 - 3N^2d^2 + 6N^2d^3 - N^3d^3 + 2N^3d^4 - 3Nd + 4d^2 - 2d^4 - 1\big)^2}.$$

Hence, also in hybrid model, P collects information only if $\psi \leq \psi_N^*$ and if

$$c \leq c_N^\star \triangleq \left(\sigma^2 - 3N\psi(2A - N\psi) \right) \Gamma_h(N, d).$$

In Fig. 10 we plot the difference $\Gamma_h(N, d) - \Gamma_p(N, d)$ in the space $(N, b) \in [0, 1/2) \times [1, +\infty)$

The figure shows that, for α large, in the region of parameters where d and N are high, the hybrid platform has a weaker incentive to collect and disclose information than a pure intermediary.

An interesting exercise is to study how the number of varieties affects the overall incentive to gather information. To this purpose, in Fig. 11, we plot the thresholds c_N^* and $c_N^\star$ as functions of N for different parameter configurations.

For large values of d the effect of an increase in N is negative in both business models, for small values of d there is an inverted-U shaped pattern.

Supplementary Information https://doi.org/10.1007/s40812-022-00218-0)

References

Abhishek, V., Jerath, K., & Zhang, Z. J. (2016). Agency Selling or Reselling? Channel Structures in Electronic Retailing. *Management Science, 62*(8), 2259–2280.

Acquisti, A., Taylor, C., & Wagman, L. (2016). The Economics of Privacy. *Journal of Economic Literature, 54,* 442–492.

Anderson, S. P., & Bedre-Defolie, O. (2021), " Hybrid Platform Model" , CEPR Working paper

Bergemann, D., & Bonatti, A. (2018). *Markets for Information*. Discussion Paper: Yale University and MIT.

Bergemann, D., & Bonatti, A. (2019). Markets for Information: An Introduction. *Annual Review of Economics, 11,* 85–107.

Calvano, E., Calzolari, G., Denicolo, V., & Pastorello, S. (2020). Artificial intelligence, algorithmic pricing, and collusion. *American Economic Review, 110*(10), 3267–97.

Evans, D. S. (2009). The online advertising industry, economics, evolution, and privacy. *Journal of Economic Perspectives, 23*(3), 37–60.

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

Gal-Or, E., Gal-Or, R., & Penmetsa, N. (2018). The role of user privacy concerns in shaping competition among platforms. *Information Systems Research, 29*(3), 698–722.

Goldfarb, A., & Tucker, C. (2011). Online display advertising: Targeting and obtrusiveness. *Marketing Science, 30*(3), 389–404.

Hagiu, A., Teh, T. H., & Wright, J. (2020), Should Platforms be Allowed to Sell on their Own Marketplaces? . *RAND Journal of Economics,* forthcoming.

Hagiu, A., & Wright, J. (2015). Marketplace or reseller? *Management Science, 61*(1), 184–203.

Jullien, B. (2012). Two-Sided B2B Platforms. In J. Waldfogel (Ed.), *Peitz, M.* The Oxford Handbook of the Digital Economy: Oxford University Press.

Kalvenes, J., & Basu, A. (2006). Design of robust business-to-business electronic marketplaces with guaranteed privacy. *Management Science, 52*(11), 1721–1736.

Karle, H., Peitz, M., & Reisinger, M. (2020). Segmentation versus agglomeration: Competition between platforms with competitive sellers. *Journal of Political Economy*, *128*(6), 2329–2374.

Kastl, J., Pagnozzi, M., & Piccolo, S. (2018). Selling information to competitive firms. *The RAND Journal of Economics*, *49*(1), 254–282.

Lam, W.M.W., & Xingyi, L. (2020), Data Usage and Strategic Pricing: Does Platform Entry Benet Independent Traders? , Working paper.

Madsen, E., & Nikhil, V. (2021). Insider Imitation , SSRN WP

Padilla, J., Piccolo, S., & Vasconcelos, H. (2021). Should vertically integrated platforms be mandated to share information with their rivals? *Economics Letters, 203,* 109849.

Peitz, M., & Reisinger, M. (2015). The economics of internet media. In S. P. Anderson, J. Waldfogel, & D. Strömberg (Eds.), *Handbook of media economics* (Vol. 1, pp. 445–530). North-Holland.

Shopova, R. (2021), Private labels in marketplaces , SSRN WP

Singh, N., & Vives, X., (1984). Price and quantity competition in a differentiated duopoly , *RAND Journal of Economics*, pp. 546–554.

Turow, J., King, J., Hoofnagle, C. J., Bleakley, A., & annìd Hennessy, M. (2009), Americans Reject Tailored Advertising and three Activities that Enable it . Available at SSRN 1478214.

Vives, X. (1984). Duopoly Information Equilibrium: Cournot and Bertrand. *Journal of Economic Theory, 34,* 71–94.

Vives, X. (2010). *Information and Learning in Markets: The Impact of Market Microstructure*. Princeton University Press.

Xu, H., Gupta, S., Rosson, M. B., & Carroll, J. M., (2012). Measuring Mobile Users' Concerns for Information Privacy , mimeo.