CSEF
Centre for Studies in Economics and Finance

WORKING PAPER NO. 583

On the Private and Social Value of Consumer Data in Vertically-Integrated Platform Markets

Jorge Padilla, Salvatore Piccolo and Helder Vasconcelos

October 2020

University of Naples Federico II — University of Salerno — Bocconi University, Milan

CSEF - Centre for Studies in Economics and Finance
DEPARTMENT OF ECONOMICS – UNIVERSITY OF NAPLES
80126 NAPLES - ITALY
Tel. and fax +39 081 675372 – e-mail: csef@unina.it
ISSN: 2240-9696

WORKING PAPER NO. 583

On the Private and Social Value of Consumer Data in Vertically-Integrated Platform Markets

Jorge Padilla[♣] **, Salvatore Piccolo**[♦] **and Helder Vasconcelos**[♠]

Abstract
We characterize and compare the private and social incentives to collect consumer data by a vertically-integrated online intermediary who competes with third-party sellers listed on its platform and is required by regulation to share with rivals all the information it gathers. With linear intermediation fees and price competition, the intermediary over-invests in accuracy compared to the social optimum when the intra-platform competition is sufficiently weak and when demand is not too responsive to quality. By contrast, the intermediary tends to under-invest in accuracy when the intra-platform competition is strong enough, and demand is sufficiently responsive to quality. With quantity competition, the intermediary always over-invests in accuracy. Importantly, when consumers exhibit privacy concerns, the over-investment problem worsens, whereas the under-investment problem mitigates. We also investigate the impact of alternative (non-linear) contractual arrangements.

JEL classification: D47, D85, L5, L81, M3

Keywords: Consumer Data, Competition, Information Accuracy, Platforms, Privacy, Value of Information, Vertical Integration

Acknowledgments: For useful comments and suggestions we would like to thank Michele Bisceglia. Jorge Padilla has advised for and against platforms in several antitrust competition investigations on both sides of the Atlantic. The views expressed in this paper are the sole responsibility of the authors and cannot be attributed to Compass Lexecon or its clients.

♣ Compass Lexecon.

♦ Università di Bergamo, Compass Lexecon and CSEF.

♠ Porto University and Compass Lexecon

Table of contents

1 Introduction

This paper investigates the data collection practices of leading tech giants, particularly those who operate vertically integrated business models, thus acting as umpire and player in their platforms. These companies typically take advantage of their privileged position to gather, manage, and, in some cases, resell information concerning their users to self-interested third parties, such as sellers and advertisers. Platforms' investments in consumer data collection are seen with great suspicion by policymakers, regulators, and society in general. Are these concerns justified? That is, is there a gap between the private and social value of such investments?

Arguably, platforms can increase the return on investment in the quality of their services to end-consumers and sellers (advertisers) by improving their understanding of consumers' preferences. Since platforms will generally not be able to appropriate the benefits associated with those quality enhancements fully, they will tend to under-invest in collecting the consumer data that facilitates them. However, the collection of consumer data may be driven by reasons other than enhancing service quality: acquiring accurate consumer data helps platforms and the sellers operating on them to price more efficiently and, therefore, extract greater rents. This benefits sellers but makes consumers worse off. As a result, platforms may simultaneously over-invest in data collection from a consumer viewpoint but under-invest from third-party sellers' standpoint.

In a nutshell, the investments of vertically integrated platforms in data collection generate positive spillovers on third-party sellers that trade on a platform in competition with the platform's own products and both positive and negative spillovers on end-consumers. When platforms cannot fully appropriate the returns of their data collection efforts — e.g., because they are unable to stipulate efficient contractual provisions with third-party sellers and/or they cannot extract the rents from such investments — their investments in data collection will not coincide with the social optimum. They may be excessive or insufficient depending on the very same parameters of consumer demand that platforms seek to apprehend; namely, consumers' willingness to pay for the platforms' services and the way that such valuation responds to changes in platform quality and the relative appeal of the platform's own products and those of the sellers operating in the platform.

Assessing the gap between the private and social incentives to invest in data collection is complex because of the platforms' investments in service quality and consumer data collection and the pricing decisions of platforms and sellers are all intimately linked strategic activities. Such link becomes even more intriguing when regulatory constraints, aimed at mitigating platforms' gatekeeper power, force them to share the information collected about their users with third-party rivals, thereby reducing the platforms' ability to appropriate the returns of their investment in data and quality.

In this paper, we build a tractable framework where the interplay between investments in quality and information collection is shaped by two intuitive forces: the extent of intra-platform competition and the responsiveness to quality of the demand for the goods traded in the platform. The objective is to characterize and compare the private and social incentives to collect consumer information by an online intermediary that competes with third-party sellers listed on its platform by introducing its own private label. Products traded on the platform are differentiated, and sellers compete in prices. The intermediary undertakes a costly investment in the platform's quality to increase demand for all products traded on the platform (e.g., better logistics, faster delivery, mandatory refund policies, guarantees, more efficient review protocols, etc.). Market demand is linear and features a random intercept. The platform collects information on this random component, and by law, it is forced to disclose it to the third-party sellers. The information disclosed is then used by all sellers (including the platform) to adjust the prices charged to end-users.

This setting allows us to answer the following research questions. Is mandatory disclosure enough to protect consumers and/or to maximize total welfare? When platforms are constrained to share their private information with rivals, will they over- or under-invest in the amount and/or accuracy of their customers' data that they gather? If so, what type of additional regulatory intervention is then needed to restore (or improve) efficiency?

With linear intermediation fees and price competition, we find that the accuracy of the data collected by the intermediary can be larger or lower than the level of accuracy that maximizes total welfare, depending on the intensity of intra-platform competition and the responsiveness of demand to quality. Specifically, the intermediary over-invests in accuracy compared to the social optimum when intra-platform competition is weak — i.e., when consumers perceive the products of the third-party sellers and the platform's private label as sufficiently differentiated — and when demand is not too responsive to quality improvements. By contrast, the intermediary tends to under-invest in accuracy when intra-platform competition is strong — i.e., the products traded on the platforms are close substitutes — and demand is sufficiently responsive to quality improvements.

By gathering more accurate information, the platform can only invest in quality when such investments materially increase consumers' willingness to pay: a demand expansion effect. This generates an incentive to gather more and more accurate information, which is greater, the higher the impact of quality on demand. The platform also benefits from more accurate information about consumer preferences because it may adjust its private label price to reflect consumer demand, thereby appropriating a larger share of consumer surplus. The platform's incentive to acquire more accurate information decreases as intra-platform competition intensifies since prices are then more cost reflector and less linked to demand — i.e., competition by third-party sellers limits the ability of the platform to extract consumers' surplus through target prices.

The effects on consumer surplus are as follows. First, since sellers use demand information to price, more accurate information harms consumers (holding quality constant). Second, since the platform uses the data also to target quality, more precise information benefits consumers (holding prices constant). The balance of those two effects determines whether the amount of information collected by the platform harms or benefits consumers. The socially optimal level of accuracy combines the benefit(s) and cost(s) of higher information accuracy on the firms' expected profits with the effect on consumer surplus. Hence, whether the platform over- or under-invests in accuracy depends on the relative strength of the forces illustrated above.

We extend the baseline analysis in several directions. First, we show that in contrast to price competition, with quantity competition, the platform always has an incentive to under-invest in accuracy relative to the social optimum. We also show that, when restricted to using linear intermediation fees, the platform strictly prefers to operate as a pure reseller and exclude competitors with quantity competition. Second, we show that large platforms — i.e., those featuring a higher number of sellers — have preferences relatively more aligned with the social planner than smaller platforms. Namely, when the number of sellers operating through the platform is large enough (i.e., when there is a sufficiently strong inter-brand competition) the over- or under-investment problem becomes less relevant. To isolate the effect of double marginalization, we also consider alternative contractual provisions that eliminate double mark-ups. The interesting result is that platforms with higher bargaining power vis-à-vis third-party sellers have incentives relatively more aligned with consumer surplus because they internalize more the impact of quality on the overall demand of the platform services. Importantly, we show that when consumers exhibit privacy concerns, the over-investment problem worsens, whereas the under-investment problem mitigates.

We organize the rest of the paper as follows. After reviewing the related literature, we present the model in Section 2. We characterize the market outcome and the optimal accuracy in Sections 2.1 and 2.2, respectively. We analyze the effect of accuracy on consumer surplus in Section 2.3 and the social value of information in Section 2.4. In Section 3, we present the extensions of the model. Section 4 concludes. All proofs are in the Appendix.

Related literature. Our paper is related to a traditional literature studying information sharing in oligopolistic markets (see, e.g., Vives, 1984, Gal-Or, 1985, and Raith, 1996, among many others). There are significant differences between our work and this literature. First, in these models, the information is assumed to be collected and shared (in the very same format) by an intermediary (e.g., a trade association) that merely plays a coordination role to facilitate the exchange of information between oligopolists. In our model, the intermediary can also be active in the final product market with its private label. Moreover, while quality provision has a crucial role in our analysis, in the traditional literature, it is typically neglected.

More recent and closely related literature is the one on markets for information.[1] Bergemann *et al.* (2018), Bergemann and Bonatti (2019) and Bergeman *et al.* (2020) investigate models in which a data intermediary collects raw data (either directly or indirectly) provided by consumers and then redistributes it (in its entirety or by adding an endogenous noise) amongst the firms. The firms then use this information to price discriminate and target advertising. They identify under which informational environments the data intermediary has incentives to lower the informativeness of the transmitted data by adding noise into it.[2] Other contributions wherein the information seller can restrict access to the information possessed as

[1] For a comprehensive survey on information markets, highlighting the central role played by data intermediaries, see Bergemann and Bonnatti (2019) and the references cited therein.

[2] As in our framework, in these models, while the noise will lower the value of the information to the firm, it will also increase consumers' willingness to pay.

well as sell noisier versions of that information include Admati and Pfleiderer (1986, 1990), Bergemann and Morris (2013) and Bimpikis *et al.* (2019). Bimpikis *et al.* (2019), in particular, demonstrate that the information provider's optimal selling strategy critically depends on the environment in which its customers interact with one another in the downstream market. More specifically, when information customers view their actions as strategic complements, the information provider finds it optimal to flood the market with highly precise signals. If instead, information customers see their actions as strategic substitutes, then the information provider finds it optimal to (*i*) reduce the quality of the information sold and (*ii*) strategically limit its market share by excluding a subset of customers from that sale.

The present paper contributes to this literature in two respects: first, we explicitly consider a vertically integrated intermediary that competes with third-party sellers in the product market; second, we introduce a 'public good' (quality of the platform) that the intermediary provides to these sellers and consumers. The impact of the provision of such public good on the sellers' competitive strategies and consumer surplus is then key to characterize the efficient level of information accuracy and compare it with the accuracy that platforms choose in an unregulated equilibrium.

Obviously, our work is related to the growing theoretical literature on the economics of privacy (see Acquisti, Taylor, and Wagman, 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.[3] We contribute to this literature by showing that, once a vertically integrated platform is forced by regulation to symmetrically allocate the collected information between its retail harm and all third-party sellers distributing through it, increasing consumer surplus and total welfare does not necessarily require the platform to collect less information than what it is optimal from a pure profit maximization perspective.

Finally, as mentioned before, there is a burgeoning literature studying how limiting the retail activity of online intermediaries in their B2B marketplaces affects consumer surplus and total welfare.[4] Hagiu *et al.* (2020), for example, show that an outright ban tends to benefit third-party sellers at the expense of consumer surplus or welfare (even after allowing for innovation by third-party sellers). On the normative side, they show that policies that limit the imitation of highly innovative third-party products and prevent steering of buyers to the intermediary's own products would lead to outcomes welfare superior to those induced by an outright ban. Although both our work and Hagiu *et al.* (2020) focus on the competitive and welfare effects of vertically integrated marketplaces, the two papers complement each other since our focus is on the informational aspects of dual-mode distribution rather than on 'copy-paste' and innovation strategies.

[3] 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, "[r]egardless of whether their private information is disclosed, consumers may dislike receiving ads that reflect too much knowledge about them".

[4] See also Jullien (2012).

2 The baseline model

Players and environment. Consider a platform (P) that, in addition to allowing a third part-seller (denoted by S_2) to distribute through its website, also owns (and is vertically integrated with) a retail unit (denoted by S_1) that distributes a private label in competition with the seller (we consider competition between third-party sellers in Section 3.2).

There is a representative consumer who has preferences defined by the following utility function

$$U(\cdot) \triangleq \sum_{i=1}^{2} (\theta + ba)\, q_i - \frac{1}{2}\left(\sum_{i=1}^{2} q_i\right)^2 - \frac{1}{1+d}\left[\sum_{i=1}^{2} q_i^2 - \frac{1}{2}\left(\sum_{i=1}^{2} q_i\right)^2\right] - \sum_{i=1}^{2} q_i p_i. \tag{1}$$

Equation (1) is a slightly modified version of the standard Shubik-Levitan quadratic utility function (see, e.g., Motta, 2004, Ch. 8.4.2.).[5] The parameter $d \geq 0$ captures the degree of differentiation between products: the larger d, the less differentiated products are (the more intense intra-platform competition). The parameter θ is a random variable capturing the heterogeneity of consumers' willingness to pay. The variable $a \geq 0$ is interpreted as the quality of the services offered by P to all platform users — e.g., logistic services such as speed, reliability and quality of delivering, refund policies, customer assistance, etc. Hence, a represents, *de facto*, a 'public good' that the platform provides to the third-party seller and to its end-users.

Maximizing (1) with respect to quantities and inverting the system of first-order conditions, the (direct) demand system is

$$q_i\left(p_i, p_j, a, \theta\right) = \frac{\theta + ba - p_i}{2} - \frac{d}{2}\left(p_i - \frac{p_i + p_j}{2}\right), \quad \forall i, j = 1, 2,$$

with $b \geq 0$ measuring the impact on demand of quality: the larger b, the more responsive is demand to quality improvements.

For simplicity, and without loss of insights, we assume that θ is distributed uniformly on the unit support $[0, 1]$. Moreover, we also assume that $b \leq \frac{7}{5}$ to guarantee that P and S_2 solve concave maximization problems, and that (in the equilibrium) prices and quality are non-negative for every $d \geq 0$.

Firms have linear production technologies with marginal costs of production normalized to zero without loss of generality. Providing quality is costly and, for tractability, we assume a quadratic cost function $\psi(a) \triangleq a^2/2$.

Information structure. Players are uninformed about θ, but P can gather an informative signal. The information policy is '*all-or-nothing*' (see, e.g., Vives, 1984, and Gal-Or, 1985). Specifically, P collects a signal $s \in [0, 1]$ that is fully informative of the true demand 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

[5] Our results generalize to alternative linear demand specifications like Singh-Vives (1984).

and independently distributed. Hence, $\mathbb{E}[\theta] = \mathbb{E}[s]$ and the conditional expectation of θ given s is linear in η — i.e.,

$$\mathbb{E}[\theta|s] = s\eta + (1-\eta)\,\mathbb{E}[\theta],$$

which is increasing in η if and only if $s \geq \mathbb{E}[\theta]$. In other words, $\mathbb{E}[\theta|s]$ exhibits increasing differences in s and η.

We assume that, because of regulation preventing discriminatory practices, the platform and the seller must have the same information — i.e., P cannot obfuscate or manipulate the information disclosed to S_2. Following Vives (2010), the cost of gathering information with precision η is convex and equal to

$$c(\eta) \triangleq \frac{c\eta}{1-\eta},$$

so that choosing a fully informative information policy ($\eta = 1$) is prohibitively costly (i.e., the Inada condition holds at $\eta = 1$). We assume that c is strictly positive but not too large in order to rule out the trivial solution in which P is always uninformed irrespective of the other parameters of the model.

Contracting, timing and equilibrium concept. Following the literature (e.g., Hagiu and Wright, 2015 and 2020) we assume that P charges S a (linear) intermediation fee $f \geq 0$ paid for every transaction made through the platform. As we shall see, linear pricing implies double marginalization — i.e., S_2 makes positive profits by passing on higher fees to consumers.[6] In Section 3.2 we consider more general contracts that eliminate double marginalization.

Following the literature on markets for data (e.g., Bergemann et al. 2018, Bergemann and Bonatti, 2019, Kastl *et al.*, 2018) we assume that P can commit to an information policy (accuracy) η.[7]

The timing of the game is as follows:

- P commits to an information policy (accuracy) η.
- θ realizes and P observes signal s, which is then disclosed to S_2.
- P sets the intermediation fee f.
- P chooses the platform's quality a and posts a price p_1 in competition with S_2 that (simultaneously) posts its price p_2.
- Demand allocates, profits materialize and fees are collected by P.

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

[6] This feature is coherent with more sophisticated theories based on asymmetric information (see, e.g., Laffont and Martimort, 2002). In adverse selection models, for example, the double marginalization problem typically arises when S_2 is privately informed about its marginal cost of production (see, e.g., Martimort and Piccolo, 2007).

[7] 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.

2.1 Market outcome

Following a backward induction logic, consider a subgame such that P has chosen accuracy η, has offered a fee f, and has observed and shared with S_2 signal $s \in [0,1]$. Let the equilibrium of this subgame be denoted by $a^*(s,f)$, $p_1^*(s,f)$, and $p_2^*(s,f)$.

S_2 solves the following maximization problem

$$\max_{p_2 \geq 0} \mathbb{E}\left[q_2\left(p_2, p_1^*(\cdot), a^*(\cdot), \theta\right)|s\right](p_2 - f).$$

The first-order condition (necessary and sufficient for a maximum) with respect to p_2 is

$$\frac{\mathbb{E}[\theta|s] + ba^*(\cdot) - p_2}{2} - \frac{d}{2}\left(p_2 - \frac{p_1^*(\cdot) + p_2}{2}\right) - \frac{2+d}{4}(p_2 - f) = 0. \tag{2}$$

Condition (2) reflects two standard effects. First, when S_2 increases its price p_2, it earns a higher profit on the infra-marginal units. Second, a higher p_2 also reduces demand for S_2's product, which means a lower sales volume.

Being vertically integrated, P and S_1 maximize their joint profit and solve the following problem

$$\max_{a \geq 0, p_1 \geq 0} \underbrace{\mathbb{E}\left[q_1\left(p_1, p_2^*(\cdot), a, \theta\right)|s\right]p_1 - \psi(a)}_{\text{Sales profit}} + \underbrace{f\mathbb{E}\left[q_2\left(p_2^*(\cdot), p_1, a, \theta\right)|s\right]}_{\text{Intermediation revenue}},$$

whose first-order conditions with respect to p_1 and a are, respectively,

$$\frac{\mathbb{E}[\theta|s] + ba - p_1}{2} - \frac{d}{2}\left(p_1 - \frac{p_1 + p_2^*(\cdot)}{2}\right) - \frac{2+d}{4}p_1 + f\frac{d}{4} = 0, \tag{3}$$

and

$$b\frac{p_1 + f}{2} - a = 0. \tag{4}$$

Condition (3) has the same intuition as (2) except for the last term. Specifically, since prices are strategic complements, a higher p_1 boosts S_2's demand which, in turn, benefits P since it collects a higher intermediation revenue when the rival sells more. Condition (4) reflects a simple trade-off: in addition to being costly, quality improvements increase the overall demand for the products distributed through the platform, which enables P to earn a higher profit from the sales of the private label and to collect higher fees also.

The solution of the system of first-order conditions (2)-(4) yields the equilibrium of the market (sub)game $p_1^*(s,f)$, $p_2^*(s,f)$ and $a^*(s,f)$. It can be shown that $p_2^*(s,f)$ is increasing in f since the intermediation fee is equivalent to S_2's marginal cost.

Hence, when choosing f, the platform solves the following maximization problem

$$\max_{f \in \Re} \mathbb{E}\left[q_1\left(p_1^*(\cdot), p_2^*(\cdot), a^*(\cdot), \theta\right)|s\right]p_1^*(\cdot) - \psi\left(a^*(\cdot)\right) + f\mathbb{E}\left[q_2\left(p_2^*(\cdot), p_1^*(\cdot), a^*(\cdot), \theta\right)|s\right].$$

By the envelope theorem, the first-order condition is

$$\underbrace{\frac{d}{4}\frac{\partial p_2^*(\cdot)}{\partial f}p_1^*(\cdot)}_{>0}+\underbrace{\mathbb{E}\left[q_2\left(p_2^*(\cdot),p_1^*(\cdot),a^*(\cdot),\theta\right)|s\right]}_{>0}-\underbrace{f\frac{2+d}{4}\frac{\partial p_2^*(\cdot)}{\partial f}}_{>0}=0. \tag{5}$$

The first term in this equation captures the strategic effect on the demand for P's private label associated with a higher intermediation fee: by inducing the rival to be less aggressive, a higher f increases the demand for P's private label. The second term reflects the fact that, for given S_2's sales, P collects a higher intermediation revenue if f is high. Finally, the third term is the expression of a second strategic effect according to which P's intermediation revenue drops when S_2 increases its price in response to a higher fee — i.e., the double marginalization effect.

Let

$$\Delta(b,d)\triangleq(6d+8)\,b^4-\left(88d+9d^2(5+d)+64\right)b^2+2(2+d)(32+d(32+9d)).$$

Solving the system of first-order conditions stated above we can characterize the equilibrium of the game including the intermediation fee.

Proposition 1 *For every signal s and accuracy η, the equilibrium market outcome is such that*

(i) P charges S_2 a fee

$$f^*(s)\triangleq\frac{(4+3d)\left(3d(4+d)+4\left(4-b^2\right)\right)}{\Delta(b,d)}\mathbb{E}[\theta|s]>0,$$

posts a price

$$p_1^*(s)\triangleq\frac{(4+3d)(8+3d)(2+d)}{\Delta(b,d)}\mathbb{E}[\theta|s],$$

and chooses a quality

$$a^*(s)\triangleq\frac{(4+3d)\left(d(13+3d)+2\left(8-b^2\right)\right)}{\Delta(b,d)}\mathbb{E}[\theta|s].$$

(ii) S_2 posts a price

$$p_2^*(s)\triangleq p_1^*(s)+\underbrace{\frac{3d(4+d)+4\left(4-b^2\right)}{\Delta(b,d)}\mathbb{E}[\theta|s]}_{\text{Double marginalization effect }(+)}>f^*(s).$$

(iii) The platform's private label is cheaper than the third-party seller's product — i.e., $p_1^*(s)<p_2^*(s)$ for every s.

(iv) All equilibrium variables are increasing in s and increasing in η if and only if $s\geq\mathbb{E}[\theta]$.

To raise an intermediation revenue, in equilibrium, P charges a positive fee to S_2. This fee is then (partially) passed on by S_2 to consumers via a higher retail price, thereby creating a standard double marginalization problem. As a result, P's private label is cheaper than S_2's product.

2.2 Optimal accuracy

We can now turn to characterize the optimal accuracy of the information collected and disclosed by P. It can be shown (see the Appendix) that P's expected profit is

$$\pi^* (\eta) \triangleq \Gamma (b, d) \times \int_0^1 \mathbb{E} [\theta|s]^2 \, ds - c (\eta) , \tag{6}$$

with

$$\Gamma (b, d) = \frac{(4 + 3d) \left(d (11 + 3d) + 2 \left(6 - b^2\right)\right)}{8\Delta (b, d)},$$

being the positive function, increasing in b and d, and independent from η plotted in Figure 1.

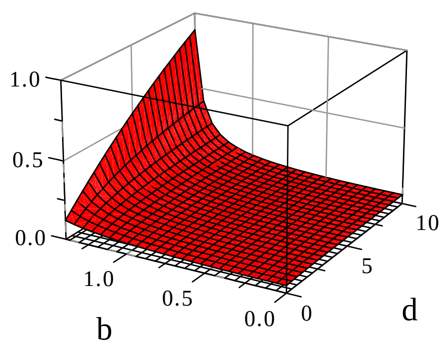


Figure 1.

The separable structure of P's expected profit in equation (6) is standard in games with quadratic payoffs or with linear demand functions (see, e.g., Vives, 2010). The function $\Gamma (b, d)$ is increasing in b because as the responsiveness of demand to quality improvements grows large, the platform invests more in quality, which shifts up the overall demand for the platform's products, thereby increasing the sales profit, the intermediation revenue and most importantly the marginal return of investing in accuracy. The reason why $\Gamma (b, d)$ is increasing in d is that as products become less differentiated — i.e., if consumers perceive products as closer substitutes — price competition restrains S_2's ability to mark up the fee charged by P, whereby mitigating the double marginalization problem and increasing P's expected profit and the marginal return of investing in accuracy.

Instead, the function

$$\int_0^1 \mathbb{E} [\theta|s]^2 \, ds = \frac{3 + \eta^2}{12}$$

captures P's return of investing in information accuracy. This function is increasing in η: as the signal structure becomes more informative, P invests more in quality when $s > \mathbb{E}[\theta]$ in order to spur demand when consumers' willingness to pay is high, thereby charging higher prices. By contrast, P invests less in quality when $s < \mathbb{E}[\theta]$ in order to save on the investment cost required to provide better quality.

Hence, we can state the following:

Proposition 2 *The level of accuracy η^* that maximizes P's expected profit is always positive and is the highest solution of*

$$\frac{2\eta(1-\eta)^2\Gamma(b,d)}{3} = c. \tag{7}$$

Moreover, η^* is increasing in b and d and decreasing in c.

Interestingly, with linear pricing, intensified competition cannot harm P because if it would do so, P would just set f equal to the monopoly price and undercut this price to exclude the rival and earn the monopoly profit.[8]

2.3 Consumer surplus

We can now study the effect of information accuracy on consumer surplus. Substituting the first-order condition (2) into the demand function and rearranging terms, we can define

$$q_i^*(\theta,s) \triangleq \frac{\theta}{2} + b^2\frac{p_1^*(s)+f^*(s)}{4} - \frac{(1+d)\,p_i^*(s)}{2} + d\frac{p_1^*(s)+p_2^*(s)}{4}, \quad \forall i=1,2.$$

Using the utility function in (1), for given demand state θ and signal s, consumer surplus can be defined as follows

$$CS(\theta,s|\eta) \triangleq \sum_{i=1}^{2}\left[\theta + b^2\frac{p_1^*(s)+f^*(s)}{2}\right] q_i^*(\theta,s) - \frac{\left[\sum_{i=1}^{2} q_i^*(\theta,s)\right]^2}{2} + \\ - \frac{(1+d)\left[p_2^*(s)-p_1^*(s)\right]^2}{2} - \sum_{i=1}^{2} q_i^*(\theta,s)\,p_i^*(s),$$

Hence, since we have assumed an 'all-or-nothing' information structure, expected consumer surplus is

$$CS(\eta) \triangleq \eta\int_0^1 CS(\theta, s=\theta|\eta)d\theta + (1-\eta)\int_0^1\int_0^1 CS(\theta,s|\eta)dsd\theta.$$

Differentiating $CS(\eta)$ with respect to η, we can show the following.

[8] In fact, it can be shown that

$$\lim_{d\to+\infty}\Gamma(b,d) = \frac{1}{8(2-b^2)} > 0.$$

Proposition 3 *There exists a function* $\Phi(b,d)$ *such that*

$$CS'(\eta) \triangleq \eta\Phi(b,d),$$

with $\Phi(b,d) \leq 0$ *if* b *and* d *are not too large, and* $\Phi(b,d) > 0$ *otherwise. Hence, consumers prefer* P *to be uninformed if* b *and* d *are not too large, otherwise they prefer* P *to be fully informed.*

Two forces shape the effect of accuracy on consumer surplus. First, since both sellers (i.e., the coalition P-S_1 and S_2) use demand data to adjust their prices, more accurate information harms consumers (holding quality constant). Second, since P uses the information also to target its quality investment, more precise information leads to greater investment and benefits consumers (keeping prices equal). The trade-off between these forces depends on the extent of intra-platform competition and the responsiveness of demand to quality improvements — i.e., on the relative size of the parameters b and d. Figure 2 shows the function $\Phi(b,d)$ in the relevant space for b and d.

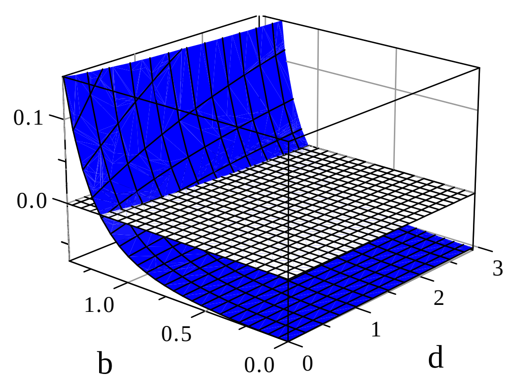


Figure 2

The figure shows that when products are sufficiently differentiated (d not too high) and quality has a low or moderate effect on demand (b not too high), consumers prefer an uninformative signal structure because this minimizes the scope for rent extraction through better targeted prices by both firms. By contrast, when d and b are both sufficiently large, consumers prefer maximal accuracy because they value relatively more P's investment in quality (more likely when b is high), and rent extraction is less of a concern because it is mitigated by intra-platform competition (high d).

2.4 The social value of information

The analysis developed above suggests that while the platform has always an incentive to gather an informative signal, consumers prefer P not to do so in the region of parameters where b and d take moderate or low values. Hence, it is interesting to examine how the tension between these effects is reflected by the level of accuracy that maximizes total welfare — i.e., consumer surplus plus P's and S_2's profits.

Consider a state of nature (θ, s), total welfare is

$$TW(\theta, s|\eta) \triangleq \sum_{i=1}^{2} \left[\theta + b^2 \frac{p_1^*(s) + f^*(s)}{2}\right] q_i^*(\theta, s) - \frac{\left[\sum_{i=1}^{2} q_i^*(\theta, s)\right]^2}{2} + \\ - \frac{(1+d)\left[p_2^*(s) - p_2^*(s)\right]^2}{2} - \psi(a^*(s)),$$

which accounts for the accuracy and quality cost. Expected total welfare is

$$TW(\eta) = \eta \int_0^1 TW(\theta, s = \theta|\eta) d\theta + (1-\eta) \int_0^1 \int_0^1 TW(\theta, s|\eta) d\theta ds - c(\eta).$$

Differentiating $TW(\eta)$ with respect to η, we can show the following.

Proposition 4 *The socially optimal level of accuracy η^{**} is positive (and lower than 1) if b and b are sufficiently large. In this region of parameters there exists a function $\Lambda(b, d) > 0$ such that η^{**} is the highest solution of*

$$\frac{2\eta(1-\eta)^2}{3} \Lambda(b, d) = c,$$

Otherwise, $\eta^{**} = 0$. Moreover, η^{**} is increasing in b and d and decreasing in c.

The socially efficient level of accuracy combines the benefit of greater accuracy on the sellers' (i.e., P-S_1 and S_2) expected profits with the effects on consumer surplus. It is increasing in b and d, since the positive impact of an increase in accuracy on consumer surplus and P's expected profit is increasing in both parameters, and is decreasing in c. In order to compare the socially optimal accuracy with the equilibrium accuracy characterized in Proposition 3, in Figure 3 we plot the function $\Lambda(b, d)$ (blue surface) and the function $\Gamma(b, d)$ (red surface) in the relevant space for b and d. The figure shows that the function $\Lambda(b, d)$ — which reflects the marginal social benefit of greater accuracy — is lower than $\Gamma(b, d)$ — which reflects P's marginal private benefit of greater accuracy — in the region of parameters where d and b are both not too large.

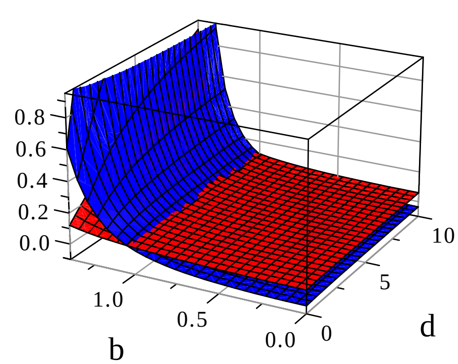


Figure 3.

Hence, we can state the following:

Proposition 5 *There exists a positive function $b(d)$, with $b'(d) < 0$ and $b(d) < 1$, such that P over-invests in accuracy compared to the socially optimal level ($\eta^* > \eta^{**}$) if and only if $b \leq b(d)$. Otherwise, P under-invests in accuracy ($\eta^* \leq \eta^{**}$).*

The intuition for this result is as follows. When choosing the optimal level of accuracy, P does not internalize the effect of this choice on consumer surplus. When d and b are not too large, quality has a moderate or weak effect on consumer surplus and the double marginalization problem is severe because intra-platform competition is not intense enough to mitigate it. Hence, from a social welfare point of view, it is efficient to reduce accuracy compared to what P would prefer in order to mitigate the adverse effect of price discrimination on consumers. As a result, in the region of parameters where $\Gamma(b,d) > \Lambda(b,d)$ an optimal regulation should induce P to lower the level of accuracy — e.g., for example by imposing restrictions on the type and quality of information that P can collect. By contrast, when b and d are sufficiently large, quality improvements benefit consumers relatively more and intra-platform competition is sufficiently strong. Hence, while choosing quality more accurately has a positive and strong impact on consumer surplus, intense intra-platform competition restrains the effect of more accurate information on the sellers' ability to extract rents from high willingness to pay consumers. However, because P does not internalize these effects, it chooses to little accuracy compared to what a social planner would prefer. Hence, an optimal regulation should actually stimulate information accuracy investments — e.g., by reducing the cost of collecting information or by subsidizing investments in quality, for example, through tailored tax subsidies.

3 Extensions and further remarks

In this section we extend the results of the baseline model in several dimensions and discuss further issues that have not been addressed in the previous analysis.

3.1 The role of double marginalization

In the baseline model, we have assumed a linear intermediation fee. The crucial implication of this hypothesis is that the market equilibrium features double marginalization. Therefore, the platform's incentive to acquire more accurate information increases as intra-platform competition intensifies because, by exerting downward pressure on prices, competition mitigates the double marginalization effect. Thus, it makes it more efficient to acquire information to target quality and price more accurately.

Would this positive relationship still be valid with alternative contractual provisions that solve the double marginalization problem? In other words, would the effects highlighted above be still valid when P can fully or partly internalize the impact of information accuracy on the profit of the third-party seller? To address this issue in the remainder of the section, we consider fixed intermediation fees as well as efficient ex-post bargaining. We will see that in both cases, new interesting comparative statics results emerge.

Fixed intermediation fees. Suppose that, instead of charging a linear intermediation fee, P 'perfectly monetizes' its brokerage function by charging a fixed fee ($F \geq 0$) that S_2 must pay before demand uncertainty realizes in order to be able to distribute through the platform. We assume that $b < 2$ in order to obtain positive prices and quality for every $d \geq 0$.

Since the fixed intermediation fee is paid ex-ante, there is no double marginalization and the first-order conditions of P's and S_2's maximization problems are the same as in the baseline model when f is set to 0. It is easy to show that in equilibrium both sellers charge the same price

$$p_F^* (s) = \frac{2}{4 + d - b^2} \mathbb{E} [\theta | s] ,$$

while P chooses quality

$$a_F^* (s) = \frac{b}{4 + d - b^2} \mathbb{E} [\theta | s] .$$

Hence, we can show the following

Proposition 6 *With a fixed intermediation fee, P's optimal accuracy η^* is the highest solution of*

$$\frac{2\eta (1 - \eta)^2}{3} \Gamma_F (b, d) = c,$$

with

$$\Gamma_F (b, d) \triangleq \frac{4 (2 + d) - b^2}{6 (4 + d - b^2)^2} > 0.$$

Moreover, η^* is increasing in b and decreasing in d and c.

When double marginalization is not an issue, the accuracy that maximizes P's profit is decreasing in d. The reason is that as products become closer substitutes, competition drives prices and profits down for all firms, which implies that P has a relatively lower incentive to gather accurate information.

We can then study consumer surplus and welfare.

Proposition 7 *Consumer surplus is increasing in η if and only if $b \leq \sqrt{2}$. The socially optimal level of accuracy η^{**} is positive (and lower than 1) and is the highest solution of*

$$\frac{2\eta (1 - \eta)^2}{3} \Lambda_F (b, d) = c,$$

with

$$\Lambda_F (b, d) \triangleq \frac{(2d + 7) b^2 - b^4 - 4}{8 (4 + d - b^2)^2} \geq 0.$$

Moreover, η^{**} is increasing in b and decreasing in d and c. Hence, there exists a function $b_F(d)$, with $b'_F(d) < 0$, such that P over-invests in accuracy if $b \leq b_F(d)$, and under-invests otherwise.

This proposition highlights two interesting results, which are worth explaining. First, consumers benefit from higher accuracy only when b is low. The reason is that with a fixed fee, P does not internalize the effect of its quality choice on S_2's demand, and thus 'under-provides' quality. The higher is b, the more expensive products will be, even though quality adjusts only partly to these higher prices (since P does not internalize the effect of a on $q_2(\cdot)$). Second, the socially optimal accuracy is decreasing in d: from a total welfare point of view, it is inefficient to invest resources in acquiring too precise information because for d large firms' profits are bound to be low because of competition, and consumer surplus is already high.

Efficient bargaining. We now also study the case of bargaining between P and S, and posit that for every realization s, P appropriates a fraction $h \in [0,1]$ of S's profit. Hence, h measures P's relative negotiation power *vis-à-vis* S.[9] For simplicity and without loss of insights, we normalize $d = 1$.

With ex post bargaining, S_2's maximization problem is essentially as before — i.e., the first-order condition is the same as in the baseline model with $f = 0$. For given p_2, P's maximization problem is instead

$$\max_{a\geq 0, p_1 \geq 0} \mathbb{E}\left[q_1(\cdot)|s\right] p_1 - \psi(a) + h\mathbb{E}\left[q_2(\cdot)|s\right] p_2,$$

whose first-order conditions with respect to p_1 and a are, respectively,

$$\frac{\mathbb{E}\left[\theta|s\right] + ba - p_1}{2} - \frac{1}{2}\left(p_1 - \frac{p_1 + p_2}{2}\right) - \frac{3}{4}p_1 + \frac{h}{4}p_2 = 0,$$

and

$$\frac{b}{2}\left(p_1 + hp_2\right) - a = 0.$$

The first condition above has an interpretation similar to (3): since prices are strategic complements, with ex-post bargaining a higher p_1 also boosts S's profit which, in turn, benefits P given that it extracts a share h of this profit (a cross-demand effect). The second condition reflects the same 'public good' effect described before with the important difference that now P's incentive to improve the platform's quality depends on S_2's price: the higher is the price that S_2 is expected to charge in equilibrium, the larger will be P's equilibrium investment in quality.

Solving the first-order conditions, the market outcome can be described as follows

$$p^*_{1,B}(s) = \frac{7+h}{35 - h - b^2(7+8h)}\mathbb{E}\left[\theta|s\right] \geq p^*_{2,B}(s) = \frac{7}{35 - h - b^2(7+8h)}\mathbb{E}\left[\theta|s\right],$$

[9] In practice, this can be seen as P charging an ad-valorem fee to S. See, e.g., Hagiu and Wright (2018) for a theory of the optimality of ad-valorem contracts in vertical contracts.

and

$$a_B^*(s) = \frac{b(7+8h)}{2(35-h-b^2(7+8h))}\mathbb{E}[\theta|s].$$

As intuition suggests, since P and S_2 bargain ex-post, P posts a higher price in equilibrium in order to relax competition and extract a larger profit from S_2. Clearly, the equilibrium quality is increasing in h: (other things being equal) the larger is the share of S_2's profit that P appropriates, the more it will invest in quality because it appropriates a larger share of the positive externality generated by its public good. As intuition suggests, when $h = 0$ (i.e., P has no bargaining power *vis-à-vis* S_2) the equilibrium prices are symmetric.

We can thus state the following.

Proposition 8 *With ex-post bargaining, P's optimal accuracy is increasing in b and h. Consumers prefer full accuracy if and only if h and b are large, and they prefer an uninformative signal otherwise. Furthermore, there exits a U-shaped function $\hat{b}_B(h)$, with $\hat{b}_B(0) > \hat{b}_B(1)$, such that P over-invests in accuracy relative to the social optimum if $b \leq \hat{b}_B(h)$, and under-invests otherwise.*

As before, P under-investments in accuracy when b is sufficiently large. However, the interesting result here is that P's incentive to under-invest in accuracy (relative to the social optimum) is non-monotone in h, and is highest when this parameter takes intermediate values. There are two effects at play. On the one hand, as h increases consumers prefer higher accuracy because P's investment in quality increases with its bargaining power. Therefore, other things being equal, consumers benefit from a more precise targeting of such quality. On the other hand, as h grows, P's private label becomes more expensive than S_2's product because P internalizes the effect of its price on the rival's profit. As a result, *ceteris paribus*, for larger values of h, P has a relatively weaker incentive than consumers to increase accuracy because a lower accuracy mitigates S_2's pricing advantage. Clearly, since $p_{1,B}^*(s) = p_{2,B}^*(s)$ at $h = 0$ and this difference is increasing in h, the latter effect is weaker for smaller values of h, whereas it is large for high values of h.

This explains the U-shaped pattern of $\hat{b}_B(h)$. Yet, overall, since $\hat{b}_B(0) > \hat{b}_B(1)$ a platform with a very high bargaining power is more likely to under-invest in accuracy than a platform with a very low bargaining power.

3.2 Multiple sellers

So far, we have assumed that there is only one seller listed on the platform. We now extend the baseline model by considering multiple competing third-party sellers distributing through the platform. To obtain tractable solutions and to focus on the effect of the number of product varieties on the social optimal accuracy, we normalize again $d = 1$ without loss of insights.

Consider a representative consumer with the following Shubick-Levitan utility function

$$U(\cdot) \triangleq \sum_{i=1}^{N} (\theta + ba)\, q_i - \frac{1}{2}\left(\sum_{i=1}^{N} q_i\right)^2 - \frac{N}{4}\left[\sum_{i=1}^{N} q_i^2 - \frac{1}{N}\left(\sum_{i=1}^{N} q_i\right)^2\right] - \sum_{i=1}^{N} q_i p_i. \tag{8}$$

As before, assume that S_i supplies product i. The corresponding demand system is

$$q_i(\cdot) = \frac{\theta + ba - p_i}{N} - \frac{1}{N}\left(p_i - \frac{1}{N}\sum_{j=1}^{N} p_j\right) \quad \forall i = 1, .., N.$$

Hence, the number N of products (varieties) traded through the platform measures the degree of intra-platform competition.

In contrast with the traditional vertical contracting literature (see, e.g., Hart and Tirole, 1990, O'Brien and Shaffer, 1992, McAfee and Schwartz, 1994, among many others) in which bilateral contracts between suppliers and retailers are typically secret, we assume that P charges sellers the same intermediation fee $f \geq 0$. This is for two reasons: first, regulatory constraints typically impose non discriminatory rules that prevent P from setting different fees to identical sellers; second, platforms usually commit to these fees up-front and do not renegotiate them. The rest of the assumptions are as in the baseline model.

We consider a 'semi-symmetric' equilibrium in which all $N-1$ third-party sellers post the same price $p_N^*(s,f)$ in state s and following an offer f, while P chooses quality $a_N^*(s,f)$ and posts a price $p_{1,N}^*(s,f)$. The analysis is essentially the same as before. Hence, P's maximization problem is

$$\max_{p_1 \geq 0, a \geq 0} \left\{ \left(\frac{\theta + ba - p_1}{N} - \frac{1}{N}\left(p_1 - \frac{p_1 + (N-1)\, p_N^*(s,f)}{N}\right)\right) p_1 - \frac{a^2}{2} + \right.$$
$$\left. + f(N-1)\left(\frac{\theta + ba - p_N^*(s,f)}{N} - \frac{1}{N}\left(p_N^*(s,f) - \frac{p_1 + (N-1)p_N^*(s,f)}{N}\right)\right)\right\},$$

whose first-order condition evaluated at equilibrium are

$$\frac{\theta + ba^*(s,f) - p_1}{N} - \frac{d}{N}\left(p_1 - \frac{p_1^*(s,f) + (N-1)\, p^*(s,f)}{N}\right) - \frac{2N-1}{N^2} p_1^*(s,f) + f\frac{N-1}{N^2} = 0, \tag{9}$$

and

$$\frac{\beta p_1^*(s,f)}{N} - a^*(s,f) + bf\frac{N-1}{N} = 0. \tag{10}$$

The intuition for these conditions is as in the baseline model, with the following differences. First, as N grows large P cares less about the intermediation revenue in its pricing decision — i.e., the last term in (9) is decreasing in N. The reason is that as intra-platform competition increases (N grows) the impact of P's private label price on the rivals demand becomes gradually smaller. Second, as N grows P cares relatively more about the quality spillover — i.e., the last term in (10) is increasing in N. This is because as the number of products increases, the positive spillover created by P's investment in quality on the rivals' demand, and thus on the intermediation revenue, becomes gradually stronger. Clearly, both effects point in the direction of increasing consumer surplus.

Each third-party seller maximizes

$$\max_{p_i \geq 0} \left(\frac{\theta + ba_N^* (s,f) - p_i}{N} - \frac{1}{N} \left(p_i - \frac{p_{1,N}^* (s,f) + p_i + (N-2)\, p_N^* (s,f)}{N} \right) \right) p_i,$$

whose first-order condition, imposing symmetry is

$$\frac{\theta + ba_N^* (s,f) - p_N^* (s,f)}{N} - \frac{1}{N} \left(p_N^* (s,f) - \frac{p_{1,N}^* (s,f) + (N-1)\, p_N^* (s,f)}{N} \right) - \frac{2N-1}{N^2} \left(p_N^* (s,f) - f \right) = 0. \tag{11}$$

Conditions (9)-(11) determine the outcome $p_N^* (s,f)$, $p_{1,N}^* (s,f)$ and $a_N^* (s,f)$. Then maximizing P's expected profit with respect to f, by the envelope theorem we have

$$\frac{N-1}{N^2} p_{1,N}^* (s,f) \frac{\partial p_N^* (s,f)}{\partial f} - f \frac{N^2-1}{N^2} \frac{\partial p_N^* (s,f)}{\partial f} + \\ + (N-1) \left(\frac{\theta + ba_N^* (s,f) - p_N^* (s,f)}{N} - \frac{1}{N} \left(p_N^* (s,f) - \frac{p_{1,N}^* (s,f) + (N-1) p_N^* (s,f)}{N} \right) \right) = 0.$$

The intuition of this condition is again as in the baseline model. The only difference worth mentioning is that as N grows the strategic effects become gradually negligible. The optimal intermediation fee is

$$f_N^* (s) = \frac{(4N-1) \left(\left(2b^2 + 9 \right) N - 12N^2 - 1 \right)}{2\Delta_N (b,N)} \mathbb{E} [\theta | s] \geq 0,$$

with

$$\Delta_N (b,N) \triangleq 1 - (4N-1)\, b^4 + (1 + N (6N+5)(2N-1))\, b^2 - 12N + 47N^2 - 48N^3.$$

Substituting into (9)-(11) and solving the system again we get

$$p_{1,N}^* (s) = \frac{2N (N-2)\, b^2 - N (13N - 10) + 1}{4\Delta_N (b,N)} \mathbb{E} [\theta | s],$$

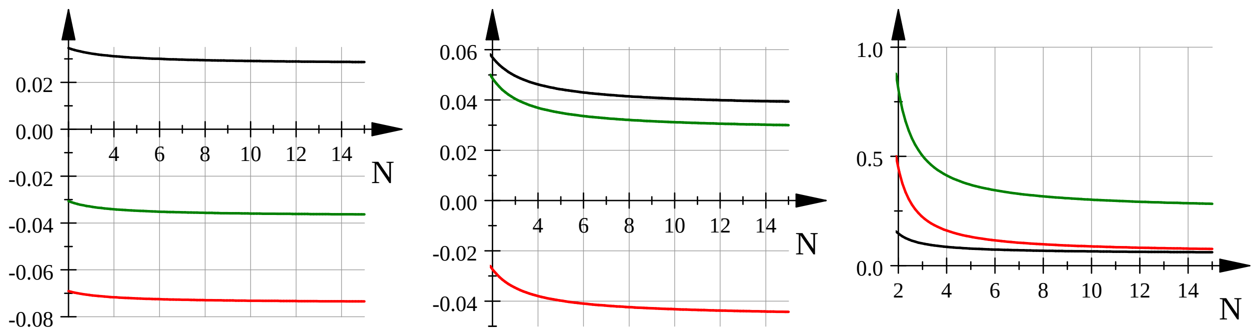
$$a_N^* (s) = \frac{b (4N-1) \left(4N + b^2 - 6N^2 \right)}{2\Delta_N (b,N)} \mathbb{E} [\theta | s] \geq 0$$

and

$$p_N^* (s) = f_N^* (s) + \underbrace{\frac{N \left(1 + 4 \left(2 + b^2 \right) N^2 - \left(7 + 6b^2 \right) N + b^2 \right)}{2\Delta_N (b,N)} \mathbb{E} [\theta | s]}_{\text{Mark-up } > 0} \geq p_N^* (s).$$

Hence, the equilibrium features again double marginalization and P's private label is cheaper than the competing products available on the platform. As intuition suggests, all these variables are increasing in b and decreasing in N.

To study the effect of N on P's incentive to under- or over-invest compared to the social optimum, in what follows we simulate the functions $\Gamma_N(b,N)$ (black line) $\Phi_N(b,N)$ (red line) and $\Lambda_N(b,N)$ (green line) for alternative values of b. As we show in the Appendix these functions play exactly the same role as in the baseline model — i.e., the measure the marginal benefit of greater accuracy for the platform, consumers and total welfare, respectively.



Panel A: b=0 Panel B: b=1 Panel C: b=$\frac{7}{5}$

Figure 4.

As intuition suggests, and in line with the previous analysis, P tends to over-invest in accuracy as b decreases. The reason is that for b small consumers do not value sufficiently quality, and are thus only harmed by more accurate information because this information only strengthens price discrimination. In Panel A where quality plays no role ($b = 0$), consumer surplus and social welfare maximization require an uninformative signal, whereas P chooses a positive accuracy. In Panel B where quality has a moderate impact on demand ($b = 1$) consumers still prefer an uninformative signal but the social optimum requires positive accuracy because more precise information benefits P relatively more than what it hurts consumers, yet in this case there is over-investment. Finally, in Panel C the effect of quality on demand is so strong ($b = \frac{7}{5}$) to induce an under-investment result. In all cases the gap between the profit maximizing accuracy and the social optimum is decreasing in N. Hence, platforms hosting a large number of sellers are likely to make choices more aligned with total welfare maximization and consumer surplus.

3.3 Quantity competition

In the baseline model and its extensions we focused on price competition. Consider now quantity competition — e.g., instances in which firms compete by setting capacities rather than prices. Consider the following (inverse) piecewise demand function

$$P(\theta, a, Q) \triangleq \max\{0, \theta + ba - dQ\},$$

with $Q \triangleq q_1 + q_2$. As before, we assume that θ is a random variable uniformly distributed on the unit support and $b \geq 0$ measures the marginal effect of quality. The parameter $d \geq 0$ is now an inverse measure of the responsiveness of demand to price — i.e., the higher d, the less responsive demand to price. We assume that $2d > b^2$ to guarantee positive quantities, price and quality.

The analysis, which is developed in the Appendix, yields a rather intuitive result.

Proposition 9 *With quantity competition, P always excludes S_2 — i.e., it sets*

$$f^*(s) = \frac{d}{2d - b^2} \mathbb{E}[\theta|s],$$

which induces $q_2^*(s) = 0$ for every s. Hence, only the private label is distributed on the market. P chooses quality

$$a^*(s) = \frac{b}{2d - b^2} \mathbb{E}[\theta|s],$$

and produces

$$q_1^*(s) = \frac{1}{2d - b^2} \mathbb{E}[\theta|s].$$

The reason why, with quantity competition, P always finds it optimal to exclude S_2 is obvious: the presence of a rival in the marketplace creates a negative externality by lowering the market price, and more so when P chooses a high quality. Hence, the optimal strategy is to set a fee high enough to deter entry. This, *de facto*, monopolizes the marketplace, and induces the platform to give up its intermediation role and operate a pure reseller (i.e., sourcing its own products and selling them, competing with outside sellers). Notice that this result does not depend on the linear structure of the contract between P and S_2: with quantity competition, P would always have an incentive to exclude rivals in order to avoid the downward pressure on the market price associated with one, or even multiple entrants.

We can now compute the optimal accuracy. It is easy to show that for every signal s, P's expected profit is

$$\pi^*(\eta) \triangleq \frac{1}{2(2d - b^2)} \times \int_0^1 \mathbb{E}[\theta|s]^2 \, ds - c(\eta),$$

which is decreasing in d: when demand is less elastic to price, P is less concerned with targeting prices, and has therefore a lower incentive to acquire accurate information. As seen before, with price competition, P's expected profit is also increasing in b: the higher the effect of quality improvements on demand, the larger P's profit.

We can thus state the following.

Proposition 10 *With quantity competition, the level of accuracy η^* that maximizes P's expected profit is always positive (and lower than 1) and is the highest solution of*

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

Moreover, η^* is decreasing in d and c and increasing in b.

Finally, we can study consumer surplus and total welfare. With a linear demand, consumer surplus can be easily computed in a quantity setting game — i.e.,

$$CS(\eta) \triangleq \frac{d}{2}\int_0^1 q_1^*(s)^2\,ds = \frac{d}{2(2d-b^2)^2} \times \int_0^1 \mathbb{E}[\theta|s]^2\,ds.$$

Hence, we can state the following.

Proposition 11 *With quantity competition, consumers surplus is always increasing in η. Hence, P always under-invests in accuracy relative to the socially optimal level of accuracy irrespective of b, d and c.*

The intuition for this result hinges on the fact that consumers surplus is quadratic in P's equilibrium quantity — i.e., consumers are risk lovers. Hence, they always benefit from higher accuracy because as η increases, supply becomes more responsive to s, and thus more uncertain. Since P always has an incentive to monopolize the downstream market, this result does not depend on the number of potential sellers.

Impossibility to exclude. The result highlighted above holds if the intermediary can freely exclude rivals, or preempt them from trading on its platform. We now assume that exclusion is forbidden. To do so, we consider a contractual environment where P charges a fixed fee F to the $N-1$ third-party sellers. The demand function is

$$P(\theta, a, Q) \triangleq \max\{0, \theta + ba - dQ\},$$

with $Q \triangleq \sum_{i=1}^N$. We assume that $2d > b^2$ to guarantee positive quantities and prices for every N.

The equilibrium outcome of the game in this case is symmetric, as seen with price competition. Specifically, each seller produces

$$q_N^*(s) = q_N^*(s) = \frac{1}{d(N+1)-b^2}\mathbb{E}[\theta|s].$$

and P chooses a quality

$$a_N^*(s) = \frac{b}{d(N+1)-b^2}\mathbb{E}[\theta|s],$$

Hence, P's profit is

$$\pi^*(\eta) \triangleq \underbrace{\frac{2Nd-b^2}{2(d(N+1)-b^2)^2}}_{\triangleq \hat{\Gamma}(b,d,N)} \times \int_0^1 \mathbb{E}[\theta|s]^2\,ds - c(\eta).$$

whereas consumer surplus is

$$CS_N(\eta) \triangleq \frac{dN^2}{2}\int_0^1 q_N^*(s)^2\,ds = \frac{dN^2}{2(d(N+1)-b^2)^2}\int_0^1 \mathbb{E}[\theta|s]^2\,ds,$$

which is still convex in the quantity. Hence, consumers still prefer maximal accuracy. We can then show the following

Proposition 12 *With ex ante contracting and no exclusion, it is still the case that with quantity competition P under-invests in accuracy compared to the social optimum. P's optimal information accuracy is increasing in b and decreasing in d and N. The under-investment problem becomes weaker for N large.*

The intuition of the result is exactly as in monopoly. Since consumers are risk lovers, they prefer maximal information accuracy, but this effect is not internalized by P. This result is standard in linear Cournot models and holds more generally even with efficient bargaining. The result that the under-investment problem is weaker for a large level of N is due to the fact that in highly competitive environments (i.e., N large) consumer surplus is already large. Therefore, the beneficial effect of information accuracy has a gradually lower impact on consumer surplus as N grows large.

3.4 Privacy concerns

nally, up to now, we have not modeled privacy concerns. Yet, the accuracy of the platform's information may directly impact consumers and reduce demand– e.g., because when P collects and shares with rivals more accurate information, consumers are more exposed to frauds, malware, cookies, etc. Consider, for example, a representative consumer with the following utility function

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

whose last term represents the (representative) consumer's privacy concerns. Clearly, the extent to which consumer personal data can be misused increases with the precision η of this information, and the volume of transactions users make through the marketplace.[10] The parameter $\psi \geq 0$ measures the significance of this damage.[11]

Demand functions are then

$$q_i\left(p_i, p_j, a, \theta\right) = \frac{\theta - \eta\psi + ba - p_i}{2} - \frac{d}{2}\left(p_i - \frac{p_i + p_j}{2}\right), \quad \forall i, j = 1, 2.$$

Notice that the intercept component of the expected demand is now

$$\mathbb{E}\left[\theta|s\right] - \eta\psi = \mathbb{E}\left[\theta\right] + \eta\left(s - \mathbb{E}\left[\theta\right] - \psi\right),$$

[10] 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.

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

meaning that the condition for information accuracy to expand demand is now tighter than before since it also reflects consumers' privacy concerns — i.e., $\mathbb{E}[\theta|s]$ increases with η if and only if $s \geq \mathbb{E}[\theta] + \psi$. Of course, demand decreases with the significance of the privacy concerns ψ.

Consider then the same game studied in the baseline model, with the addition of privacy concerns that reduce the demand intercept in proportion to the information accuracy. For simplicity, and without loss insights, assume $d = 1$. Following the previous backward induction logic, we can characterize the equilibrium of the game, which features again double marginalization and a lower price for P's private label.[12] That is:

$$p_1^*(s) = \frac{231}{2(219 - 103b^2 + 7b^4)}(\mathbb{E}[\theta|s] - \eta\psi) \leq p_2^*(s) = \frac{293 - 8b^2}{2(219 - 103b^2 + 7b^4)}(\mathbb{E}[\theta|s] - \eta\psi),$$

$$a^*(s) = \frac{7b(4-b)(b+4)}{219 - 103b^2 + 7b^4 + 219}(\mathbb{E}[\theta|s] - \eta\psi),$$

and

$$f^*(s) = \frac{7(31 - 4b^2)}{2(219 - 103b^2 + 7b^4)}(\mathbb{E}[\theta|s] - \eta\psi) < p_2^*(s).$$

As expected, all these values are decreasing in the privacy concerns, as measured by the parameter ψ that we assume not too large to avoid uninteresting corner solutions.

In order to study the effects of privacy concerns on the private and social incentives to acquire consumer data, in Figure 5 we simulate the private and social marginal returns on information accuracy. In other words, we simulate the following derivatives (see the online Appendix):

$$\frac{\partial[\pi^*(\eta) + c(\eta)]}{\partial\eta},$$

and

$$\frac{\partial[TW(\eta) + c(\eta)]}{\partial\eta},$$

which are plotted as a function of $\eta \in [0, 1]$.

The black lines in either panel represent the marginal social return on the investment in accuracy, while the red lines in either panel represent P's private marginal return on the investment in accuracy. Solid lines, instead, represent the private and social return on the investment accuracy when there are no privacy concerns ($\psi = 0$); dashed lines correspond to the case with privacy concerns ($\psi = 0.1$).[13]

In Panel A, absent privacy concerns, P over-invests in accuracy compared to the social optimum since the solid red line lies above the solid black one. In this scenario, privacy concerns make the over-investment problem more severe since the gap between the red dashed line, and the black shed line expands. By contrast, in Panel B, absent privacy concerns, P under-invests in accuracy relative to the social optimum:

[12] The first-order conditions are the same as in the baseline model except for the change in the demand intercept that now accounts for the privacy concerns term $\eta\psi$.

[13] Results are robust to alternative specification.

since the solid red line lies above the solid black line. In this scenario, the introduction of privacy concerns makes the under-investment problem less severe since the gap between the red dashed line and the black dashed line drops.

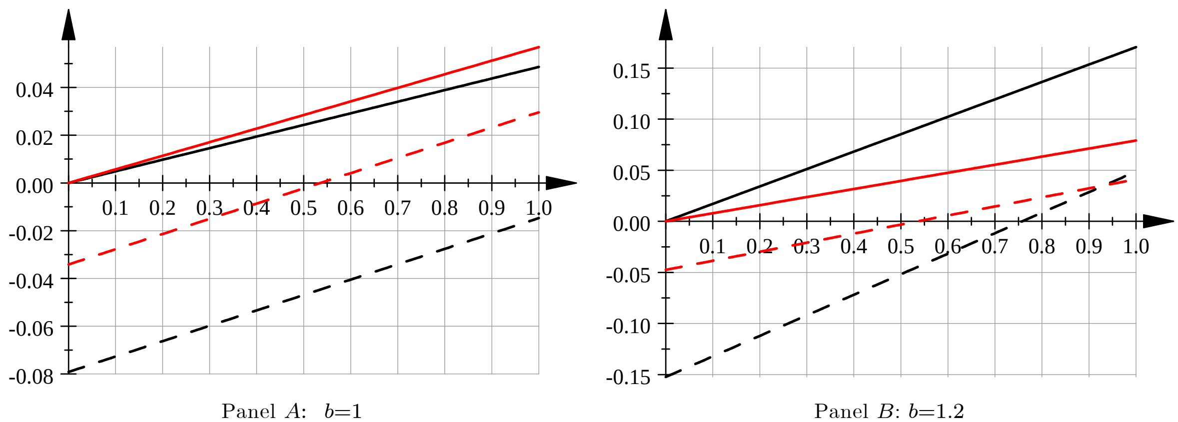


Figure 5

Besides stimulating the investment in quality and facilitating rent extraction, with privacy concerns, increased accuracy also directly reduces consumer surplus and demand, which hurts both the social and private marginal value of information. Yet, while the demand reduction impacts profits and welfare in the same manner (because total welfare accounts for profits), the negative impact on consumer surplus is only present in total welfare. As a result, other things being equal, privacy concerns tend to make the over-investment problem relatively more severe than the under-investment problem.

4 Takeaway and concluding remarks

The use or misuse of consumer data in platform markets currently is at the heart of the policy debate on the benefits and costs of digitalization. Specifically, policy makers and pundits have expressed concerns that platforms may leverage their gatekeeper role to extract personal information that can be used to predict consumers' willingness to pay and set targeted prices. Others have noted that such information may also be used to offer better-targeted complementary services and products, thus warning regulators against excessive privacy protection. By developing a tractable framework to account for the complex incentives of vertically integrated platforms, this paper has shown that the accuracy of the data collected by these intermediaries is not necessarily excessive from a total welfare perspective.

In our model, the intermediary over-invests in accuracy compared to the social optimum when the intra-platform competition is weak — i.e., when consumers perceive the products of the third-party sellers and the platform's private label as sufficiently differentiated — and when demand is not too responsive to the quality of the platform. Regulations protecting privacy — e.g., restraining the type and quality of information collected and shared by platforms — may constitute appropriate public policy. By contrast, the intermediary tends to under-invest in accuracy when the intra-platform competition is strong —

i.e., the products traded on the platforms are close substitutes — and demand is sufficiently responsive to quality. In this case, regulations that facilitate rather than hinder data collection would be most appropriate.

In sum, our analysis suggests that the only sensible way of approaching privacy regulation in digital markets is a case-by-case approach, which takes account of the business model at stake, the peculiar competitive and market conditions under which a platform operates.

References

[1] Abhishek, V., Jerath, K., and Zhang, Z.J., (2016), "Agency Selling or Reselling? Channel Structures in Electronic Retailing," *Management Science*, Vol. 62(8), pp. 2259-2280.

[2] Acquisti, A., C. Taylor, and Wagman, L., (2016), "The Economics of Privacy", *Journal of Economic Literature*, 54, pp. 442-492.

[3] Admati, A. R., and Pfleiderer, P., (1986), "A monopolistic market for information", *Journal of Economic Theory*, 39(2), pp. 400-438.

[4] Admati, A. R., and Pfleiderer, P., (1990), "Direct and indirect sale of information", *Econometrica*, 58(4), pp. 901-928.

[5] Armstrong, M., (2006), "Competition in Two-sided Markets", *RAND Journal of Economics*, Vol. 37, No. 3, pp. 668–691.

[6] Armstrong, M., (2007), "Two-Sided Markets: Economic Theory and Policy Implications". In: Choi, J. P. (Ed.), Recent Developments in Antitrust, MIT Press: Cambridge.

[7] Bergemann, D., Bonatti, A., and Smolin, A., (2018), "The design and price of information," *American Economic Review,* 108, pp. 1-48.

[8] Bergemann, D., and Bonatti, A., (2019), "Markets for Information: An Introduction", *Annual Review of Economics*, Vol. 11, pp. 85-107.

[9] Bergemann, D., Bonatti, A., and Gan, T., (2020) "The economics of social data." mimeo, MIT.

[10] Bergemann, D., and S. Morris (2013), "Robust Predictions in Games with Incomplete Information", *Econometrica*, 81, pp. 1251-1308.

[11] Bimpikis, K., D. Crapis, and A. Tahbaz-Salehi (2019), "Information Sale and Competition", *Management Science*, Vol. 65, Issue 6, pp. 2646-2664.

[12] Caillaud, B. and Jullien, B., (2003), "Chicken & Egg: Competition Among Intermediation Service Providers", *RAND Journal of Economics*, Vol. 34, No. 2, pp. 309-328.

[13] Evans, D.(2003) "The Antitrust Economics of Multi-Sided Platform Markets", *Yale Journal on Regulation*, Vol. 20, pp. 325-381.

[14] Evans, David S., (2009), "The Online Advertising Industry, Economics, Evolution, and Privacy.", *Journal of Economic Perspectives,* 23 (3): 37–60.

[15] Gal-Or, E., (1985), "Information sharing in oligopoly", *Econometrica*, 53(2), pp. 329-343.

[16] Goldfarb, A., & Tucker, C., (2011), "Online display advertising: Targeting and obtrusiveness. Marketing Science, 30(3), pp. 389-404.

[17] Hagiu. A, and Wright, J., (2015), "Marketplace or Reseller?", *Management Science*, 61(1), pp. 184–203.

[18] Hagiu. A, Jullien, B. , and Wright, J., (2020), "Creating platforms by hosting rivals", *Management Science*, forthcoming.

[19] Hagiu. A, Teh, T-H , and Wright, J., (2020), "Should Amazon be allowed to sell on its own marketplace?", mimeo.

[20] Hart, O., and Tirole, J. (1990). "Vertical Integration and Market Foreclosure." *Brookings Papers on Economic Activity. Microeconomics*, 205-286.

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

[22] Kalvenes, J., and Basu, A., (2006), "Design of Robust Business-to-Business Electronic Marketplaces with Guaranteed Privacy", *Management Science*, 52(11): 1721-1736.

[23] Kastl, J., Pagnozzi, M., and Piccolo S., (2018) "Selling information to competitive firms" *The RAND Journal of Economics,* Vol. 49, No. 1, pp. 254-282.

[24] Laffont, J. J., and Martimort, D., (2002). *The theory of incentives: the principal-agent model.* Princeton university press.

[25] Martimort, D., and Piccolo, S., (2007) "Resale price maintenance under asymmetric information", *International Journal of Industrial Organization,* 25(2): 315-339.

[26] McAfee, R. P., and Schwartz, M. (1994). "Opportunism in Multilateral Vertical Contracting: Nondiscrimination, Exclusivity, and Uniformity." *American Economic Review*, 84: 210-230.

[27] Motta, Massimo. *Competition policy: theory and practice.* Cambridge University Press, 2004.

[28] O'Brien, D. P., and Shaffer, G. (1992). "Vertical Control with Bilateral Contracts." *RAND Journal of Economics*, 23: 299-308.

[29] Peitz, M., and Reisinger, M., (2015), "The economics of internet media," Handbook of media economics. Vol. 1. North-Holland, 445-530.

[30] Raith, M. (1996), "A General Model of Information Sharing in Oligopoly", *Journal of Economic Theory*, 71: 260-288.

[31] Rochet, J.-C. and Tirole, J., (2003), "Platform Competition in Two-Sided Markets", *Journal of the European Economic Association*, Vol. 1: 990–1029.

[32] Rochet, J.-C. and Tirole, J., (2006), "Two-Sided Markets: A Progress Report", *RAND Journal of Economics*, Vol. 37(3): 645-667.

[33] Tucker, C.E., (2014), "Social networks, personalized advertising and privacy controls", Journal of Marketing Research, Vol. 51(5): 546–562.

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

[35] Singh, N., and Vives, X., (1984). "Price and quantity competition in a differentiated duopoly", *RAND Journal of Economics*, 546-554.

[36] Tucker, C.E., (2014), "Social networks, personalized advertising and privacy controls", Journal of Marketing Research, Vol. 51(5), pp. 546–562.

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

[38] Vives, X. (1984), "Duopoly Information Equilibrium: Cournot and Bertrand", *Journal of Economic Theory*, 34: 71-94.

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

[40] Xu, H., Gupta, S., Rosson, M. B., and Carroll, J. M., (2012). "Measuring Mobile Users' Concerns for Information Privacy", mimeo.

Appendix

Second-order conditions. Differentiating again (4)-(3) we obtain the following Hessian Matrix

$$\mathcal{H}(b,d) \triangleq \begin{bmatrix} -\frac{2+d}{2} & \frac{b}{2} \\ \frac{b}{2} & -1 \end{bmatrix},$$

which is negative semi-definite if and only if

$$\frac{2+d}{2} \geq \frac{b^2}{4} \quad \Leftrightarrow \quad 4+2d \geq b^2,$$

which is always satisfied in the region of parameters under consideration.

Proof of Proposition 1. Solving the first-order conditions (2)-(3) we obtain

$$p_1^*(s,f) = \frac{2}{4+d-b^2}\mathbb{E}[\theta|s] + \frac{(4+3d)b^2+3d(2+d)}{(3d+4)(4+d-b^2)}f, \tag{12}$$

$$p_2^*(s,f) = \frac{2}{4+d-b^2}\mathbb{E}[\theta|s] + \frac{8+(3d+2)b^2+d(3d+8)}{(4+3d)(4+d-b^2)}f, \tag{13}$$

and

$$a^*(s,f) = \frac{b}{4+d-b^2}\mathbb{E}[\theta|s] + \frac{b(3d+8)(d+1)}{(3d+4)(4+d-b^2)}f. \tag{14}$$

Substituting these expressions into (5) we obtain

$$f^*(s) \triangleq \frac{(4+3d)\left(3d(4+d)+4\left(4-b^2\right)\right)}{\Delta(b,d)}\mathbb{E}[\theta|s].$$

Substituting $f^*(s)$ back into (12), (13) and (14) and rearranging we obtain the equilibrium outcome.

Notice that $p_1^*(s) \geq 0$ if and only if

$$\Delta(b,d) \triangleq (6d+8)b^4 - \left(88d+9d^2(5+d)+64\right)b^2 + 2(2+d)(32+d(32+9d)) \geq 0.$$

$\Delta(b,d)$ is a biquadratic in b with $\Delta(b=0,d)>0$ and $\Delta(b,d)\leq 0$ for

$$b \leq \bar{b}(d) \triangleq \sqrt{\frac{64+88d+45d^2+9d^3-(8+3d)\sqrt{d(d+1)(33d+9d^2+32)}}{4(4+3d)}}.$$

It can be shown that $\bar{b}(d)$ is decreasing in d and that $\lim_{d\to+\infty}\bar{b}(d) \approx \frac{7}{5}$. Hence, to guarantee $\Delta(b,d)\geq 0$ for every $d \geq 0$ we assume $b \leq \frac{7}{5}$. ■

Proof of Proposition 2. From the first-order condition (4) we obtain that

$$\pi^*(s) \triangleq \frac{2+d}{4}p_1^*(s)^2 - \frac{d}{4}f^*(s)p_1^*(s) + f^*(s)\mathbb{E}[q_2(p_2^*(s),p_1^*(s),a^*(s),\theta)|s] - \frac{a^*(s)^2}{2}$$

Substituting $p_1^*(s)$, $a^*(s)$, $p_2^*(s)$ and $f^*(s)$ into $\pi^*(s)$ and taking expectations we obtain $\pi^*(\eta)$ — i.e.,

$$\pi^*(\eta) = \Gamma(b,d)\int_0^1 [\eta s + (1-\eta)\mathbb{E}[\theta]]^2 ds - c(\eta) = \frac{3+\eta^2}{3}\Gamma(b,d) - c(\eta).$$

Differentiating with respect to η and rearranging we obtain the first-order condition

$$\eta(1-\eta)^2 = \frac{3c}{2\Gamma(b,d)}.$$

For c not too large, this condition has two solutions in η. However, it can be checked that only the highest one corresponds to a maximum (since the function $\eta(1-\eta)^2$ is inverted-U shaped). The comparative statics with respect to b, d and c follows immediately from the fact that $\Gamma(b,d)$ is increasing in b and d as shown in Figure 1. ■

Proof of Proposition 3. Substituting the equilibrium outcomes into $CS(\theta, s|\eta)$, taking the expectation and differentiating we have

$$CS'(\eta) = \eta\Phi(b,d),$$

with

$$\begin{aligned}\Phi(b,d) = & -\frac{(2+d)\left(17152d+17392d^2+8952d^3+2331d^4+243d^5+6912\right)}{12\Delta(b,d)^2} + \\ & + \frac{(2+d)\left(5040d+5072d^2+2649d^3+720d^4+81d^5+2080\right)}{3\Delta(b,d)^2}b^2 + \\ & - \frac{15760d+17\,216d^2+10\,416d^3+3789d^4+810d^5+81d^6+6112}{12\Delta(b,d)^2}b^4 + \\ & + \frac{(3d+4)\left(64+88d+45d^2+9d^3\right)}{3\Delta(b,d)^2}b^6 - \frac{(3d+4)^2}{3\Delta(b,d)^2}b^8.\end{aligned}$$

It can be shown that $\Phi(0,d) < 0$ and $\Phi(7/5,d) > 0$. Hence, by the mean-value theorem there must exist a function $b(d)$ such that $\Phi(b,d) \geq 0$ for every $b \geq b(d)$ and $\Phi(b,d) < 0$ otherwise. Figure 2 shows that this function is unique and that $b'(d) < 0$. ■

Proof of Proposition 4. Substituting the equilibrium outcomes into $TW(\theta, s|\eta)$, taking the expectation with respect to θ and s and differentiating the resulting expected welfare we have

$$TW'(\eta) = \frac{2\eta}{3}\Lambda(b,d) - \frac{c}{(1-\eta)^2},$$

with

$$\begin{aligned}\Lambda(b,d) &= -\frac{(2+d)\left(7936d+7632d^2+3672d^3+873d^4+81d^5+3328\right)}{8\Delta(b,d)^2}+\\&+\frac{4(d+2)\left(5344d+5432d^2+2820d^3+747d^4+81d^5+2144\right)}{8\Delta(b,d)^2}b^2+\\&-\frac{15888d+17264d^2+10416d^3+3789d^4+810d^5+81d^6+6176}{8\Delta(b,d)^2}b^4+\\&+\frac{4(4+3d)\left(88d+45d^2+9d^3+64\right)}{8\Delta(b,d)^2}b^6-\frac{4(3d+4)^2}{8\Delta(b,d)^2}b^8.\end{aligned}$$

Hence, in an interior solution, the first-order condition is

$$\eta(1-\eta)^2 = \frac{3c}{2\Lambda(b,d)}.$$

As seen in the proof of Proposition 2, for c not too large, this condition has two solutions in η and that only the highest one corresponds to a maximum (since the function $\eta(1-\eta)^2$ is inverted-U shaped).

The function $\Lambda(b,d)$ is illustrated in Figure 6 below.

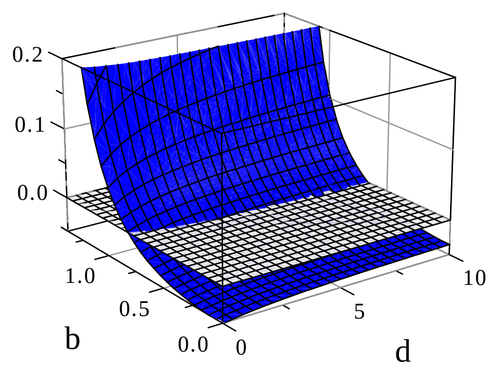


Figure 6.

It can be easily seen that $\Lambda(0,d) < 0$ at $b = 0$ and that $\Lambda(7/5,d) > 0$. Hence, Figure 6 implies that $\eta^{**} > 0$ only if b and d are sufficiently large, and that η^{**} is increasing in b and d. ■

Proof of Proposition 5. The proof of this result is immediately implied by Figure 3. ■

Proof of Proposition 6. The first-order condition for S_2's maximization problem is

$$\frac{\mathbb{E}[\theta|s]+ba-p_2}{2}-\frac{d}{2}\left(p_2-\frac{p_1+p_2}{2}\right)-\frac{2+d}{4}p_2=0.$$

The first-order conditions with respect to p_1 and a for P's maximization problem are

$$\frac{\mathbb{E}[\theta|s]+ba-p_1}{2}-\frac{d}{2}\left(p_1-\frac{p_1+p_2}{2}\right)-\frac{2+d}{4}p_1=0,$$

$$b\frac{p_1}{2} - a = 0.$$

These three conditions yield the market outcome

$$p_1^*(s) = p_2^*(s) = p_F^*(s) = \frac{2}{4+d-b^2}\mathbb{E}[\theta|s],$$

$$a_F^*(s) = \frac{b}{4+d-b^2}\mathbb{E}[\theta|s].$$

Next, by the first-order conditions we have

$$\begin{aligned}
\pi_1^F(s) &= \frac{2+d}{2}p_F^*(s)^2 - \frac{b^2}{8}p_F^*(s)^2 \\
&= p_F^*(s)^2\left(\frac{4(2+d)-b^2}{8}\right) \\
&= \underbrace{\frac{4(2+d)-b^2}{2(4+d-b^2)^2}}_{=\Gamma_F(b,d)}\mathbb{E}[\theta|s]^2
\end{aligned}$$

Integrating with respect to s and netting out the cost of gathering information we have

$$\pi_F(\eta) = \frac{4(2+d)-b^2}{2(4+d-b^2)^2}\int_0^1 (\eta s + (1-\eta)\mathbb{E}[\theta])^2 ds - c(\eta)$$

$$= \frac{2(2+d)-b^2}{24(4+d-b^2)^2}\left(3+\eta^2\right) - c(\eta).$$

Differentiating with respect to η and rearranging, in an interior solution

$$\frac{2\eta\left(1-\eta^2\right)}{3}\underbrace{\frac{2(2+d)-b^2}{8(4+d-b^2)^2}}_{=\Gamma_F(b,d)} = c, \tag{15}$$

which yields immediately the result. ■

Proof of Proposition 7. Let

$$\begin{aligned}
q(s,\theta) &\triangleq \frac{\theta + ba^*(s) - p^*(s)}{2} \\
&= \frac{\theta - \left(2-b^2\right)\mathbb{E}[\theta|s]}{2}.
\end{aligned}$$

Consumer surplus is

$$U_F(s,\theta) \triangleq 2(\theta + ba_F^*(s))\,q_F^*(\theta,s) - 2q_F^*(\theta,s)^2 - q_F^*(\theta,s)\,p_F^*(s).$$

Substituting $q_F^*(\theta,s)$, $a_F^*(s)$ and $p_F^*(s)$ we have

$$CS_F(\eta) \triangleq \eta\int_0^1 CS_F(\theta, s=\theta|\eta)d\theta + (1-\eta)\int_0^1\int_0^1 CS_F(\theta,s|\eta)dsd\theta$$

$$= -\eta^2\frac{\left(2-b^2\right)\left(6+2d-b^2\right)}{24\left(4+d-b^2\right)^2} + \frac{b^4-2\left(4+d\right)b^2+4\left(5d+d^2+7\right)}{24\left(4+d-b^2\right)^2}.$$

Differentiating with respect to η

$$CS_F'(\eta) = -\eta\frac{\left(2-b^2\right)\left(6+2d-b^2\right)}{12\left(4+d-b^2\right)^2},$$

which immediately yields the result on consumer surplus.

Consider now total welfare. It can be shown that

$$TW_F(\theta,s) = 2\left(\theta + ba_F^*(s)\right)q_F^*(\theta,s) - 2q_F^*(\theta,s)^2 - \frac{a_F^2(s)}{2},$$

Substituting $q_F^*(\theta,s)$, $a_F^*(s)$ and $p_F^*(s)$ and accounting for the cost of acquiring information we have

$$TW_F(\eta) \triangleq \eta\int_0^1 TW_F(\theta, s=\theta|\eta)d\theta + (1-\eta)\int_0^1\int_0^1 TW_F(\theta,s|\eta)dsd\theta - c(\eta)$$

$$= \frac{b^4-(2d+7)b^2-4}{24\left(4+d-b^2\right)^2}\eta^2 + \frac{1}{24}\frac{32d-2b^2d-11b^2+b^4+4d^2+52}{24\left(4+d-b^2\right)^2} - c(\eta).$$

The first-order condition with respect to η is

$$\frac{2\eta\left(1-\eta^2\right)}{3}\underbrace{\frac{(2d+7)b^2-b^4-4}{8\left(4+d-b^2\right)^2}}_{=\Lambda_F(b,d)} = c. \tag{16}$$

Comparing (16) with (15)

$$\Lambda_F(b,d) - \Gamma_F(b,d) = -\frac{b^4-2b^2(d+4)+(2d+8)}{8\left(4+d-b^2\right)^2},$$

which is positive if and only if

$$b \geq b_F(d) \triangleq \sqrt{4+d-\sqrt{6d+d^2+8}},$$

with $b_F(d) < 2$ and $b_F'(d) < 0$. The result follows immediately. ■

Proof of Proposition 8. The first-order condition for S_2's maximization problem is

$$\frac{\mathbb{E}\left[\theta|s\right]+ba-p_2}{2}-\frac{1}{2}\left(p_2-\frac{p_1+p_2}{2}\right)-\frac{3}{4}p_2=0.$$

The first-order conditions with respect to p_1 and a, respectively, for P's maximization problem are

$$\frac{\mathbb{E}\left[\theta|s\right]+ba-p_1}{2}-\frac{1}{2}\left(p_1-\frac{p_1+p_2}{2}\right)-\frac{3}{4}p_1+\frac{h}{4}p_2=0,$$

$$b\frac{p_1}{2}-a-\frac{bh}{2}p_2=0.$$

The solution of these conditions yield the market outcome

$$p_{1,B}^*\left(s\right)=\frac{7+h}{35-h-b^2\left(7+8h\right)}\mathbb{E}\left[\theta|s\right]\geq p_{2,B}^*\left(s\right)=\frac{7}{35-h-b^2\left(7+8h\right)}\mathbb{E}\left[\theta|s\right],$$

$$a_B^*\left(s\right)=\frac{b\left(7+8h\right)}{2\left(35-h-b^2\left(7+8h\right)\right)}\mathbb{E}\left[\theta|s\right].$$

Substituting into P's expected profit

$$\pi_B^*\left(\eta\right)=\left(3+\eta^2\right)\frac{294-\left(8h+7\right)^2b^2+8h\left(35-h\right)}{96\left(35-h-b^2\left(7+8h\right)\right)^2}-c\left(\eta\right),$$

Differentiating with respect to η we have

$$\underbrace{\frac{294-\left(8h+7\right)^2b^2+8h\left(35-h\right)}{12\left(35-h-b^2\left(7+8h\right)\right)^2}}_{\Gamma_B(h,b)}\eta-\frac{c}{\left(1-\eta\right)^2}.$$

Hence, in an interior maximum, the optimal accuracy solves

$$\Gamma_B\left(h,b\right)\eta\left(1-\eta\right)^2=c.$$

The function $\Gamma_B\left(h,b\right)$ is positive and increasing in b and h.

Consider consumer surplus. It can be shown that

$$CS_B'\left(\eta\right)=\eta\underbrace{\frac{2\left(7+8h\right)\left(35-h\right)b^2-\left(7+8h\right)^2b^4-\left(56+5h\right)\left(14-h\right)}{12\left(35-h-b^2\left(7+8h\right)\right)^2}}_{=\Phi(b,h)}$$

Figure 7 plots $\Phi\left(b,h\right)$ in the relevant region of parameters for h and b.

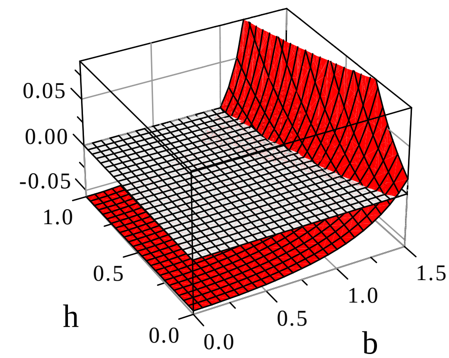


Figure 7.

Consumer surplus is maximized by a fully informative signal ($\eta = 1$) if and only if

$$b \geq \tilde{b}(h) \triangleq \sqrt{\frac{35 - h - \sqrt{3}\sqrt{147 - 2h(14 - h)}}{8h + 7}},$$

which is decreasing in h.

Consider total welfare. It can be shown that

$$TW_B'(\eta) = \underbrace{\frac{b^2(8h+7)(63-10h) - (8h+7)^2 b^4 - \left(28h + 3h^2 + 196\right)}{12\left(35 - h - b^2(7+8h)\right)^2}}_{=\Lambda_B(b,h)} \eta - \frac{c}{(1-\eta)^2}.$$

In an interior maximum the social optimal accuracy solves the following first-order condition

$$\eta(1-\eta)^2 \Lambda_B(b,h) = c.$$

Figure 8 plots the function $\Lambda_B(b,h)$

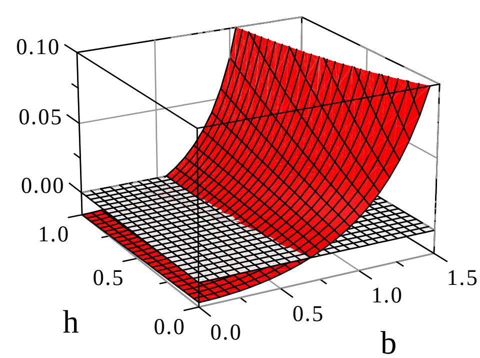


Figure 8.

As a result, total welfare maximization requires a positive accuracy if and only if

$$b \geq b_B(h) \triangleq \sqrt{\frac{63 - 10h - \sqrt{3185 - 4h(343 - 22h)}}{14 + 16h}},$$

with $b'_B(h) < 0$.

Comparing P's optimal accuracy with the social optimum level we have

$$\Lambda_B(b,h) - \Gamma_B(h,b) = \frac{b^4(8h+7)^2 - 2b^2(8h+7)(35-h) + h(308-5h) + 490}{12\left(35 - h - b^2(7+8h)\right)^2}. \tag{17}$$

Solving (17) with respect to b, it follows that P over-invests for

$$b \geq \hat{b}_B(h) \triangleq \sqrt{\frac{35 - h - \sqrt{3}\sqrt{245 - 2h(63-h)}}{8h+7}},$$

which is decreasing for moderate or low h and increasing for relatively large h. Figure 9 plots the difference $\Lambda_B(b,h) - \Gamma_B(h,b)$

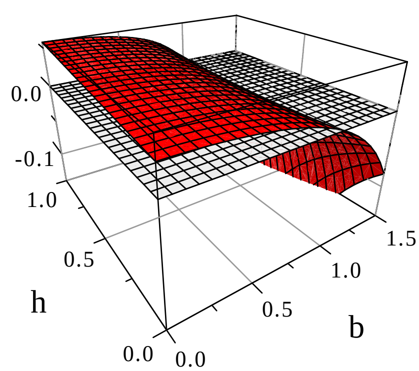


Figure 9.

The result follows immediately. ■

N **sellers**. The system of equations (9)-(11) pins down the equilibrium of the subgame

$$p^*_N(s,f) = \frac{N}{3N - \beta^2 - 1}\mathbb{E}[\theta|s] + f\frac{1 - 7N + 8N^2 + \left(4N^2 - 6N + 1\right)b^2}{(4N-1)(3N - b^2 - 1)},$$

$$p^*_{1,N}(s,f) = \frac{N}{3N - \beta^2 - 1}\mathbb{E}[\theta|s] - f(N-1)\frac{1 - 5N - b^2(4N-1)}{(4N-1)(3N - b^2 - 1)}$$

and

$$a^*_N(s,f) = \frac{b}{3N - b^2 - 1}\mathbb{E}[\theta|s] + f\frac{2b(6N-1)(N-1)}{(4N-1)(3N - b^2 - 1)}$$

Substituting these expressions into the first-order condition with respect to f we obtain

$$f^*(s) = \frac{(4N-1)\left(\left(2b^2+9\right)N - 12N^2 - 1\right)}{2\Delta_N(b,N)}\mathbb{E}[\theta|s] \geq 0,$$

with

$$\Delta_N(b,N) \triangleq 1 - (4N-1)b^4 + (1 + N(6N+5)(2N-1))b^2 + 47N^2 - 48N^3 - 12N.$$

Plugging $f^*(s)$ into (9)-(11) $p^*_{1,N}(s,f)$, $a^*_N(s,f)$ and $p^*_N(s,f)$ we obtain the market outcome.

Next, substituting the equilibrium prices and quality into P's expected profit, the derivative of this function with respect to η is

$$\eta \underbrace{\frac{(4N-1)\left(1+N+b^2-4N^2\right)}{12\Delta_N(b,N)}}_{=\Gamma_N(b,N)} - c'(\eta).$$

In Figure 4 the black lines correspond to $\Gamma_N(b,N)$ evaluated at $b=0$, $b=1$ and $b=\frac{7}{5}$.

In order to determine consumer surplus and total welfare, let

$$q^*_{1,N}(\theta,s) \triangleq \frac{\theta + ba^*_N(s) - p^*_{1,N}(s)}{N} - \frac{N-1}{N^2}\left(p^*_{1,N}(s) - p^*_N(s)\right),$$

and

$$q^*_N(\theta,s) = \frac{\theta + ba^*_N(s) - p^*_N(s)}{N} - \frac{1}{N^2}\left(p^*_N(s) - p^*_{1,N}(s)\right).$$

Consumer surplus is

$$\begin{aligned} U_N(\theta,s) \triangleq{}& (N-1)(\theta + ba^*_N(s))\,q^*_N(\theta,s) + (\theta + ba^*_N(s))\,q^*_{1.N}(\theta,s) + \\ & -\frac{1}{4}\left[q^*_{1,N}(\theta,s) + (N-1)\,q^*_N(\theta,s)\right]^2 - \frac{N}{4}\left[q^*_{1,N}(\theta,s)^2 + (N-1)\,q^*_N(\theta,s)^2\right] + \\ & - q^*_{1,N}(\theta,s)\,p^*_{1,N}(s) - (N-1)\,q^*_N(\theta,s)\,p^*_N(s) - \psi(a^*_N(s)) - c(\eta). \end{aligned}$$

Hence, expected consumer surplus is

$$CS_N(\eta) = \eta\int_0^1 U_N(\theta,\theta)d\theta + (1-\eta)\int_0^1\int_0^1 U_N(\theta,s)d\theta ds.$$

Differentiating with respect to η it can be shown that

$$CS'_N(\eta) = \eta\Phi_N(b,N).$$

Total welfare is

$$TW_N(\theta,s) \triangleq (N-1)\left(\theta + ba_N^*(s)\right) q_N^*(\theta,s) + \\ + \left(\theta + ba_N^*(s)\right) q_{1.N}^*(\theta,s) - \frac{1}{4}\left[q_{1,N}^*(\theta,s) + (N-1)\, q_N^*(\theta,s)\right]^2 + \\ - \frac{N}{4}\left[q_{1,N}^*(\theta,s)^2 + (N-1)\, q_N^*(\theta,s)^2\right] - \psi\left(a_N^*(s)\right) - c(\eta).$$

Hence, expected total welfare is

$$TW_N(\eta) = \eta \int_0^1 TW_N(\theta,\theta)d\theta + (1-\eta)\int_0^1\int_0^1 TW_N(\theta,s)d\theta ds.$$

Differentiating with respect to η it can be shown that

$$TW_N'(\eta) = \eta \Lambda_N(b,N) - c'(\eta).$$

The functions $\Phi_N(b,N)$ and $\Lambda_N(b,N)$ are simulated fin Figure 4 (see the online Appendix).

Proof of Proposition 9. Consider a subgame such that P offers f given s. Focus on an equilibrium of this subgame where S_2 produces $q_2^*(s,f)$ while P produces $q_1^*(s,f)$ and chooses quality $a^*(s,f)$. P's maximization problem is

$$\max_{a\geq 0, q_1 \geq 0} \left(\mathbb{E}[\theta|s] + ba - d(q_1+q_2)\right) q_1 - \psi(a) + fq_2.$$

The first-order conditions with respect to q_1 and a are

$$\mathbb{E}[\theta|s] + ba - 2dq_1 - dq_2 = 0,$$

$$bq_1 - a = 0.$$

S_2 maximizes

$$\max_{a\geq 0, q_1 \geq 0} \left(\mathbb{E}[\theta|s] + ba - d(q_1+q_2)\right) q_2 - fq_2.$$

The first-order condition is

$$\mathbb{E}[\theta|s] + ba - 2dq_2 - dq_1 - f = 0.$$

Hence, the equilibrium of the subgame features

$$a^*(s,f) = b\frac{\mathbb{E}[\theta|s] + f}{3d - b^2}, \quad q_1^*(s,f) = \frac{\mathbb{E}[\theta|s] + f}{3d - b^2}, \quad q_2^*(s,f) = \frac{d\left(\mathbb{E}[\theta|s] - 2f\right) + b^2 f}{d\left(3d^2 - b^2\right)}.$$

Moving backward to P's choice of f we have

$$\max_{f\geq 0} \left(\mathbb{E}[\theta|s] + ba^*(s,f) - d\left(q_1^*(s,f) + q_2^*(s,f)\right)\right) q_1^*(s,f) - \psi\left(a^*(s,f)\right) + fq_2^*(s,f).$$

The optimal intermediation fee solves

$$-d\frac{\partial q_2^*\left(s,f\right)}{\partial f}q_1^*\left(s,f\right)+q_2^*\left(s,f\right)+f\frac{\partial q_2^*\left(s,f\right)}{\partial f}=0 \quad \Rightarrow \quad f^*\left(s\right)=\frac{d}{2d-b^2}\mathbb{E}\left[\theta|s\right].$$

Hence, $q_2^*\left(s,f^*\left(s\right)\right)=0$. P's expected profit is then

$$\pi^*\left(\eta\right)\triangleq\frac{1}{2\left(2d-b^2\right)}\times\int_0^1\mathbb{E}\left[\theta|s\right]^2ds-c\left(\eta\right),$$

which is decreasing in d and increasing in b. The result then follows immediately. ■

Proof of Proposition 10. Consider an equilibrium where P produces $q_{1,N}^*\left(s\right)$ and chooses quality $a_N^*\left(s\right)$, whereas each third-party seller produces $q_N^*\left(s\right)$. P's maximization problem is

$$\max_{a\geq0,q_1\geq0}\left(\mathbb{E}\left[\theta|s\right]+ba-d\left(q_1+\left(N-1\right)q_N^*\left(s\right)\right)\right)q_1-\psi\left(a\right).$$

The first-order conditions with respect to q_1 and a are

$$\mathbb{E}\left[\theta|s\right]+ba_N^*\left(s\right)-2dq_{1,N}^*\left(s\right)-d\left(N-1\right)q_N^*\left(s\right)=0,$$

$$bq_1-a=0.$$

S_2 maximizes

$$\max_{a\geq0,q_1\geq0}\left(\mathbb{E}\left[\theta|s\right]+ba_N^*\left(s\right)-d\left(q_1^*+\left(N-2\right)q^*\left(s\right)\right)\right)q_2-fq_2,$$

whose first-order condition is

$$\mathbb{E}\left[\theta|s\right]+ba_N^*\left(s\right)-dq_{1,N}^*\left(s\right)-d\left(N-2\right)q_N^*\left(s\right)-2dq_N^*\left(s\right)=0.$$

The market outcone is then

$$a_N^*\left(s\right)=\frac{b}{d\left(N+1\right)-b^2}\mathbb{E}\left[\theta|s\right],$$

$$q_N^*\left(s\right)=q_N^*\left(s\right)=\frac{1}{d\left(N+1\right)-b^2}\mathbb{E}\left[\theta|s\right].$$

Hence, P's expected profit is

$$\underbrace{\frac{2Nd-b^2}{2\left(d\left(N+1\right)-b^2\right)^2}}_{\triangleq\hat{\Gamma}\left(b,d,N\right)}\times\int_0^1\mathbb{E}\left[\theta|s\right]^2ds-c\left(\eta\right).$$

The first, order condition is

$$\hat{\Gamma}_N\left(b,d,N\right)\frac{2\eta\left(1-\eta\right)^2}{3}=c,$$

with

$$\frac{\partial \hat{\Gamma}_N(b,d,N)}{\partial b} = b\frac{(2N-1)\,d + Nd - b^2}{\left(d\,(N+1) - b^2\right)^3} > 0,$$

$$\frac{\partial \Gamma(b,d,N)}{\partial d} = -\frac{Nd\,(N+1) - b^2}{\left(d\,(N+1) - b^2\right)^3} < 0,$$

$$\frac{\partial \hat{\Gamma}_N(b,d,N)}{\partial N} = -d^2\frac{N-1}{\left(d\,(N+1) - b^2\right)^3} < 0.$$

Consumer surplus is

$$CS(\theta, s) = \frac{\theta + ba^*(s) - (\theta + ba^*(s) - dNq^*(s))}{2} Nq^*(s) = \frac{d}{2}N^2 q^*(s)^2.$$

Hence, in expected terms

$$CS_N(\eta) = \underbrace{\frac{dN^2}{2\left(d\,(N+1) - b^2\right)^2}}_{\triangleq \hat{\Phi}(b,d,N)} \int_0^1 \mathbb{E}\left[\theta|s\right]^2 ds,$$

which is always increasing in η. Comparing $\hat{\Phi}(b,d,N)$ and $\hat{\Gamma}(b,d,N)$

$$\hat{\Phi}(b,d,N) - \hat{\Gamma}(b,d,N) = \frac{Nd\,(N-2) + b^2}{2\left(d\,(N+1) - b^2\right)^2} > 0,$$

Hence, P under-invests in accuracy relative to the efficient level of accuracy. Finally,

$$\frac{\partial\left[\hat{\Phi}(b,d,N) - \hat{\Gamma}(b,d,N)\right]}{\partial N} = d\frac{d - N\left(2d - b^2\right)}{\left(d\,(N+1) - b^2\right)^3} \leq 0 \quad \Leftrightarrow \quad N \geq \frac{d}{3d - b^2},$$

which concludes the proof. ■