Online Appendix to "The value of personal information in online markets with endogenous privacy"

Rodrigo Montes, Wilfried Sand-Zantman, and Tommaso Valletti

Appendix B: Proofs of the results in Section 5

Proof of Lemma 2

First we show that there is no equilibrium when information is sold to only one firm with certainty. By contradiction, consider a candidate equilibrium in which consumers anticipate that only firm A buys the data. In this case, the outcome of the game is the same as in Proposition 5. However, the data supplier could deviate from selling consumer information exclusively to firm A and sell it to firm B. If it does, then the DS can charge a price equivalent to the profits that B generates from old consumers. Observe that, after the DS deviates at stage 1, both firms set the same basic prices at stage 3 (as in Proposition 5); from stage 0.5 onward θ_2 is fixed (and is also defined in Proposition 5). Therefore,

$$\pi_B^O = \int_{\frac{3(t-c)}{8t}}^{1} \left(\frac{t-c}{2} + t(2\theta - 1) + t\right) d\theta = \frac{10ct - 21c^2 + 75t^2}{64t}.$$

We can easily show that π_B^O is strictly greater than the price of information given in Proposition 6. So if consumers anticipate that the information will be bought by A (resp., by B), then the data supplier has an incentive to sell to B (resp., to A). As a consequence, there cannot be an equilibrium in which the data supplier necessarily allocates data to a single firm.

Next we establish that there is no equilibrium in which information is sold to both firms with probability 1. Consider a candidate equilibrium under which that certainty prevails. In this case, as shown in Section 3, consumers do not pay the privacy cost. But then the DS would be incentivized to sell the information to only one firm (and we are back to the scenario covered in Section 3). Since consumers should rationally anticipate this deviation, it cannot occur in equilibrium.

Proof of Proposition 7

Suppose consumers anticipate that each firm will have information exclusively with probability $\frac{1}{2}$. Consider a symmetric equilibrium in which a fraction θ_A pays for privacy (to hide from A) and a fraction $1-\theta_B$ pays for privacy (to hide from B), where $\theta_A = 1-\theta_B$.

If a consumer close to A pays for privacy, then he anticipates a utility of $-p_A^a - c - t\theta_A + v$. Yet, if the same consumer does not pay for privacy, then his expected utility is

$$\frac{1}{2}(-p_A^a - t(2\theta_A - 1) - t(1-\theta_A) + v) + \frac{1}{2}(-p_B^a - t\theta_A - t(1-2\theta_A) + v).$$

The first term represents the utility derived by this consumer from paying a tailored price (as calculated before) to A; the second term represents the utility from paying a tailored price to B. Each term is multiplied by the probability of the respective event. The choice of an indifferent consumer θ_A is then

$$\theta_A = \frac{p_B^a - p_A^a + t - 2c}{2t}.$$

Given θ_A, firm A's profits can be written as

$$\frac{1}{2}\left(\theta_1(p_A, p_B)p_A + \theta_A p_A + \int_{\theta_A}^{\theta_B} [p_B + (1-2\theta)t]\, dx\right) + \frac{1}{2}(\theta_1(p_A, p_B)p_A + \theta_A p_A);$$

here $\theta_1(p_A, p_B)$ denotes A's market share among new consumers. The first term represents profits when the firm has data, and the second term when it does not (but B does). By symmetry, firms A and B solve (respectively)

$$\arg\max_{p_A} [\theta_1(p_A, p_B) + \theta_A]p_A \quad \text{and} \quad \arg\max_{p_B} [(1-\theta_1(p_A, p_B)) + \theta_A]p_B.$$

The equalities $p_B = p_B^a$ and $p_A = p_A^a$ (consumers' rational expectations) lead to equilibrium prices $p_A = p_B = 2(t-c)$ and $\theta_A = \frac{t-2c}{2t}$. Then, for such a solution to be feasible, the condition $c \leq \frac{t}{2}$ must be satisfied. In this solution, the data supplier can set a price T equal to the profits in the personalized market (i.e., $[\theta_A, \theta_B]$). This is because these profits are in addition to the profits the firm would make without data. Therefore,

$$T = \int_{\theta_A}^{\theta_B} [2(t-c) + (1-2\theta)t]\; dx = 4c - \frac{4c^2}{t}.$$

When $c > t/2$, consumers do not pay for privacy and the personalized market consists of all the old consumers. Since $\theta_A = 0$, it follows that $p_A = p_B = t$. Then the extra profits enabled by having information (equivalently, the price of that information) is:

$$T = \int_0^1 p_A(\theta)\, d\theta = t.$$

By construction of the price T, neither firm A nor firm B deviates at stage 2. Finally, it can be shown that the data broker does not deviate—by selling to *both* firms—because more revenue is always generated by selling to one firm only. Formally, such deviation (D) would yield the revenue

$$T^D = 2\int_{\theta_A}^{1/2} t(1-2\theta)d\theta < T.$$

Price deviations by A and B at stage 3 can be ruled out because, since both firms are playing their best response in the anonymous market, changing basic prices could only generate losses. Furthermore, the data supplier has no incentive to change the allocation rule because it is indifferent between selling the data to A or B—and this indifference is consistent with consumer expectations.

Proof of Lemma 3

Let us consider, without loss of generality, a possible deviation from firm A. Its profit, when it tries to poach on firm B's turf, can be written as

$$\pi_A^D = \frac{1}{t}[p_B - p_A + 2t\theta_A]p_A + \int_{\theta_A}^{\frac{1}{2}} (p_B + (1-2\theta)t)\, d\theta$$

Note that only the first term of the RHS is affected by the choice of firm A. Using the equilibrium price for firm B, $p_B = t(1+2\theta_A)$, and optimizing with respect to p_A leads to $p_A^D = \frac{t}{2}(4\theta_A + 1)$ and the optimal deviation profits are then

$$\pi_A^D = \frac{t}{4}(1+4\theta_A)^2 + \int_{\theta_A}^{\frac{1}{2}} (p_B + (1-2\theta)t)\, d\theta,$$

The equilibrium profits being equal to $\pi_A = \frac{t}{2}(1+2\theta_A)^2 + \int_{\theta_A}^{\frac{1}{2}} (p_B + (1-2\theta)t)\, d\theta$, we find that the equilibrium profits are greater when $\theta_A < \sqrt{2}/4$. At last, we use the definition θ_A, given by the consumer θ indifferent between paying $p_A + c$ and $p_B + (1-2\theta)t$, to derive $\theta_A = \frac{t-c}{2t}$ and the result of the lemma. Note at last that $\tilde{c} = \frac{t}{2}(2-\sqrt{2}) \approx 0.29t < \bar{c} \approx 0.96t$.

Proof of Lemma 4

We assume that firm A has full information about old consumers – if they do not pay the privacy cost – whereas firm B has only a signal on whether the old consumer's type is above of below σ with $\sigma \in [1/2, 1]$. B will then set 3 prices: p_B for the anonymous market; $\underline{p}_B$ for the old consumers with type less that σ; $\overline{p}_B$ for the old consumers with type greater than σ. As before, we denote by θ_A the cut-off type of old consumers that hide from A (located to the left) and $1-\theta_B$ the cut-off type of old consumers that hide from B (located to the right).

We first characterize the standard price for the anonymous market, that is p_A and p_B. The price p_A is chosen to maximize

$$p_A\theta_A + (\frac{1}{2} + \frac{p_B - p_A}{2t})p_A.$$

This leads to $p_A = \frac{p_B+t(1+2\theta_A)}{2}$. Similarly $p_B = \frac{p_A+t(1+2\theta_B)}{2}$. Combining the two equations leads to

$$p_A = t + \frac{2t}{3}[2\theta_A + \theta_B] \text{ and } p_B = t + \frac{2t}{3}[2\theta_B + \theta_A].$$

To go further, we now investigate the old consumers market. For any consumer with type $\theta \leq 1/2$, firm A can offer a price that matches any positive price set by firm B. For types $\sigma \geq \theta > 1/2$, Firm A can match B's offer if and only if $\theta \leq \frac{\underline{p}_B+t}{2t}$. Let us look at the choice of $\underline{p}_B$. One should consider two cases. If $\underline{p}_B \geq 2\sigma t - t$, then A will capture the whole market in $[1/2, \sigma]$. If $\underline{p}_B \leq 2\sigma t - t$, then B's program is $\max_{\underline{p}_B} \int_{\frac{\underline{p}_B+t}{2t}}^{\sigma} \underline{p}_B d\theta$ so $\underline{p}_B = t[\sigma - \frac{1}{2}]$.

We now look at the choice of $\bar{p}_B$. For this, we assume (and will check the consistency *ex post*) that $\sigma < 1-\theta_B$. If B can capture a share of the market $[\sigma, 1-\theta_B]$, its program is $\max_{\bar{p}_B} \int_{\frac{\bar{p}_B+t}{2t}}^{1-\theta_B} \bar{p}_B d\theta$ s.t. $\frac{\bar{p}_B+t}{2t} > \sigma$. This leads to $\bar{p}_B = t[\frac{1}{2} - \theta_B]$. We need $\frac{\bar{p}_B+t}{2t} \geq \sigma$, that is $\sigma \leq \frac{3-\theta_B}{4}$. And note that we also need $\frac{\bar{p}_B+t}{2t} < 1-\theta_B$, that is $\theta_B < 1/2$.

Next we turn to the choice of privacy. Without any protection, the price paid by the agents located on the right of the line is $\bar{p}_B$. With protection, the price paid is $p_B + c = t + \frac{2t}{3}[2\theta_B + \theta_A] + c$. Since $\bar{p}_B \leq t/2$ and $p_B + c \geq t + c$, no consumer located to the right of the Hotelling line will choose to hide. Therefore, we will have $\theta_B = 0$ and $\bar{p}_B = t/2$. Remark that, with $\theta_B = 0$ and $\bar{p}_B = t/2$, we need $\sigma \leq \frac{3}{4}$.

Let us now compute θ_A, characterized by $p_A + c = p_A(\theta_A)$. We know that $p_A(\theta) = \underline{p}_B + (1-2\theta)t$. This means that θ_A is such that

$$p_A + c = \underline{p}_B + (1-2\theta_A)t \Leftrightarrow p_A = \frac{t}{2}\left[2\sigma + 1 - 4\theta_A\right] - c.$$

Moreover, since $\theta_B = 0$, we have $p_A = t + \frac{4t}{3}\theta_A$. Using these two relationships leads to $\theta_A = \frac{3}{10}\left[\sigma - \frac{1}{2} - \frac{c}{t}\right]$. This implies that Firm A's price on the anonymous market is $p_A = \frac{2t}{5}\left[\sigma + 2 - \frac{c}{t}\right]$ and that Firm B's price on the anonymous market is $p_B = \frac{t}{5}\left[\sigma + \frac{9}{2} - \frac{c}{t}\right]$.

Let us summarize this case.

1. For $\sigma \leq 3/4$ and $c < (\sigma - 1/2)t$ (that is $\theta_A > 0$), then $\theta_A = \frac{3}{10}\left[\sigma - \frac{1}{2} - \frac{c}{t}\right]$, $\theta_B = 0$, $p_A = \frac{2t}{5}\left[\sigma + 2 - \frac{c}{t}\right]$, $p_B = \frac{t}{5}\left[\sigma + \frac{9}{2} - \frac{c}{t}\right]$, $\bar{p}_B = t/2$ and $\underline{p}_B = t[\sigma - \frac{1}{2}]$.

2. For $\sigma \leq 3/4$ and $c \geq (\sigma - 1/2)t$, we will have $\theta_A = 0$, $p_A = p_B = t$, $\bar{p}_B = t/2$ and $\underline{p}_B = t[\sigma - \frac{1}{2}]$.

Let us compute the profit in the first sub-case where $c < (\sigma - 1/2)t$:

$$\begin{aligned}
\pi_A &= \theta_A p_A + \int_0^{1/2+(p_B-p_A)/2t} p_A d\theta + \int_{\theta_A}^{\frac{\underline{p}_B+t}{2t}} \Big(\underline{p}_B + (1-2\theta)t\Big)\, d\theta + \int_{\sigma}^{\frac{\bar{p}_B+t}{2t}} (\bar{p}_B + (1-2\theta)t)\, d\theta \\
&= p_A\big[\theta_A + 1/2 + \frac{p_B - p_A}{2t}\big] + \int_{\theta_A}^{\frac{\underline{p}_B+t}{2t}} \Big(\underline{p}_B + (1-2\theta)t\Big)\, d\theta + \int_{\sigma}^{\frac{\bar{p}_B+t}{2t}} (\bar{p}_B + (1-2\theta)t)\, d\theta.
\end{aligned}$$

But we can show that $\frac{p_B - p_A}{2t} + \theta_A + 1/2 = \frac{2\theta_A}{3} + 1/2 = \frac{1}{5}\big[\sigma + 2 - \frac{c}{t}\big]$ and $\frac{\underline{p}_B+t}{2t} = \frac{\sigma+1/2}{2}$ and $\frac{\bar{p}_B+t}{2t} = 3/4$. Therefore

$$\pi_A = \frac{2t}{25}\big[\sigma + 2 - \frac{c}{t}\big]^2 + \frac{t}{100}\big[2\sigma + 4 + 3\frac{c}{t}\big]^2 + (\frac{3}{4} - \sigma)^2 t.$$

Similarly

$$\begin{aligned}
\pi_B &= \int_{1/2+\frac{p_B-p_A}{2t}}^{1} (p_B)\, d\theta + \int_{\frac{\underline{p}_B+t}{2t}}^{\sigma} \Big(\underline{p}_B\Big)\, d\theta + \int_{\frac{\bar{p}_B+t}{2t}}^{1} (\bar{p}_B)\, d\theta \\
&= \frac{t}{50}\big[\sigma + \frac{9}{2} - \frac{c}{t}\big]^2 + \frac{t}{2}\big[\sigma - \frac{1}{2}\big]^2 + \frac{t}{8}.
\end{aligned}$$

Here, we need to define and characterize the revenues of the DS. To each firm, he can threaten to sell the whole set of information to the competing firm, which would lead to a profit equal to the value π_B derived in Corollary 2. Let us denote $\underline{\pi}_B$ this minimal value. The total revenue of the DS is therefore $R = \pi_A + \pi_B - 2\underline{\pi}_B$. We can now differentiate R with respect to σ:

$$\begin{aligned}
\frac{\partial R}{\partial \sigma} &= \frac{4t}{25}\big[\sigma + 2 - \frac{c}{t}\big] + \frac{t}{25}\big[2\sigma + 4 + 3\frac{c}{t}\big] - \frac{t}{2}\big[\frac{3}{4} - \sigma\big] \\
&\quad + \frac{t}{25}\big[\sigma + \frac{9}{2} - \frac{c}{t}\big] + t\big[\sigma - \frac{1}{2}\big]
\end{aligned}$$

Note first that $\frac{\partial^2 R}{\partial \sigma^2} > 0$. Note also that $\frac{\partial R}{\partial \sigma}\,|_{\sigma=1/2} = \frac{t}{25}\big[25 - 2\frac{c}{t}\big] - \frac{t}{8}$ which is positive (since $c < t$). So, for $c < (\sigma - 1/2)t$, the optimal σ for the DS in 3/4.

We now look at the second subcase where $c \geq (\sigma - 1/2)t$. Then, Firm A's profit writes as

$$\begin{aligned}
\pi_A &= \frac{t}{2} + \int_0^{\frac{\underline{p}_B+t}{2t}} \Big(\underline{p}_B + (1-2\theta)t\Big)\, d\theta + \int_{\sigma}^{\frac{\bar{p}_B+t}{2t}} \big[\bar{p}_B + (1-2\theta)t\big] d\theta \\
&= \frac{t}{2} + \frac{t}{4}\left[\sigma + \frac{1}{2}\right]^2 + t\left[\frac{3}{4} - \sigma\right]^2.
\end{aligned}$$

Similarly, Firm B's profit are

$$\begin{aligned}
\pi_B &= \frac{t}{2} + \int_{\frac{\underline{p}_B+t}{2t}}^{\sigma} \underline{p}_B d\theta + \int_{\frac{\bar{p}_B+t}{2t}}^{1} \bar{p}_B d\theta \\
&= \frac{t}{2} + \frac{t}{2}[\sigma - \frac{1}{2}]^2 + \frac{t}{8}.
\end{aligned}$$

As before, the Data Seller's revenue is $R = \pi_A + \pi_B - 2\underline{\pi}_B$ and we differentiate R with respect to σ.

$$\frac{\partial R}{\partial \sigma} = t[\frac{5\sigma}{2} - \frac{5}{4}] + t[\sigma - \frac{1}{2}] > 0, \text{ for all } \sigma \in [1/2, 3/4].$$

Therefore, for any $c \in [0, t]$ and $\sigma \in [1/2, 3/4]$, the optimal level of precision is $\sigma = 3/4$.

We now look at the case where $\sigma \geq 3/4$. In such a case, firm A will not be able to compete with Firm B for the consumers too close to 1, i.e., for consumers with types greater than σ. Indeed, the information gained by Firm B about these captive consumers is more precise, so Firm B will set the maximal price such that $\frac{\bar{p}_B+t}{2t} = \sigma$ so $\bar{p}_B = t[2\sigma - 1]$. We now consider a consumer with type $\theta \in [\sigma, 1]$ and his choice to pay the privacy cost or not. As before, he must compare $\bar{p}_B$ and $p_B + c$. But again, $\bar{p}_B < t$ and $p_B + c \geq t + c$ so no agent to the right of the line will choose to pay the privacy cost, i.e., $\theta_B = 0$ as in the first case studied. This implies that θ_A and the prices p_A and p_B are the same as before. We now compute the firms' profit in order to characterize the DS's revenue.

As far as firm A is concerned,

$$\begin{aligned}
\pi_A &= \theta_A p_A + \int_0^{1/2+(p_B-p_A)/2t} p_A d\theta + \int_{\theta_A}^{\frac{\underline{p}_B+t}{2t}} \left(\underline{p}_B + (1-2\theta)t\right) d\theta \\
&= \frac{2t}{25}\left[\sigma + 2 - \frac{c}{t}\right]^2 + \frac{t}{100}\left[2\sigma + 4 + 3\frac{c}{t}\right]^2.
\end{aligned}$$

Regarding firm B,

$$\begin{aligned}
\pi_B &= \int_{1/2+\frac{p_B-p_A}{2t}}^{1} (p_B)\, d\theta + \int_{\frac{\underline{p}_B+t}{2t}}^{\sigma} \left(\underline{p}_B\right) d\theta + \int_{\sigma}^{1} (\bar{p}_B)\, d\theta \\
&= \frac{t}{50}\left[\sigma + \frac{9}{2} - \frac{c}{t}\right]^2 + \frac{t}{2}(\sigma - \frac{1}{2})(\frac{7}{2} - 3\sigma).
\end{aligned}$$

As before, we differentiate the DS's revenue with respect to σ. This leads to

$$\begin{aligned}
\frac{\partial R}{\partial \sigma} &= \frac{t}{25}\left[6\sigma + 12 - \frac{c}{t}\right] + \frac{t}{25}\left[\sigma + \frac{9}{2} - \frac{c}{t}\right] + \frac{t}{2}[5 - 6\sigma] \\
&= \frac{t}{50}\left[158 - 136\sigma - 4\frac{c}{t}\right] > 0, \text{ for all } \sigma \in [3/4, 1].
\end{aligned}$$

Let us assume now that $c > (\sigma - 1/2)t$, that is $\theta_A = 0$. We then get $p_A = p_B = t$, $\underline{p}_B = t(\sigma - \frac{1}{2})$ and $\bar{p}_B = 2t(\sigma - \frac{1}{2})$. Then

$$\begin{aligned}
\pi_A &= \frac{t}{2} + \int_0^{\frac{\underline{p}_B+t}{2t}} \left(\underline{p}_B + (1-2\theta)t\right) d\theta \\
&= \frac{t}{2} + \frac{t}{4}(\sigma + 1/2)^2.
\end{aligned}$$

Similarly

$$\begin{aligned}\pi_B &= \frac{t}{2}+\int_{\frac{\underline{p}_B+t}{2t}}^{\sigma}(\underline{p}_B)d\theta+\int_{\sigma}^{1}\bar{p}_B d\theta \\ &= \frac{t}{2}+\frac{t}{2}(2\sigma-1)(\frac{7}{4}-\frac{3}{2}\sigma).\end{aligned}$$

Still denoting R the DS's revenue, we differentiate R with respect to σ. This leads to

$$\frac{\partial R}{\partial \sigma}=\frac{t}{2}(\frac{11}{2}-5\sigma)>0, \text{ for all } \sigma\in[3/4,1]$$

Hence, the DS's revenue is still optimized for $\sigma=1$, which completes the proof of Lemma 4.

Proof of Proposition 8

As shown in Lemma 4, the optimal level of precision for the DS is to choose $\sigma=1$. Using the expression of π_A and π_B derived above, and setting $\sigma=1$, we can write the equilibrium profit of the DS as

$$\begin{aligned}R &= \frac{2t}{25}\left[3-\frac{c}{t}\right]^2+\frac{t}{100}\left[6+3\frac{c}{t}\right]^2+\frac{t}{50}\left[\frac{11}{2}-\frac{c}{t}\right]^2+\frac{t}{8}-2\underline{\pi}_B \\ &= \frac{t}{100}\left[\frac{337}{2}-36\frac{c}{t}+19\left(\frac{c}{t}\right)^2\right]+\frac{t}{8}-2\underline{\pi}_B\end{aligned}$$

where $\underline{\pi}_B=\frac{(5t-c)^2}{32t}$ is the profit of Firm B derived in Corollary 2. This must be compared to the DS's revenue in the benchmark case derived in Proposition 6, that is $T=\frac{9c^2-2ct+57t^2}{64t}$. Then

$$\begin{aligned}T\geq R \quad &\Leftrightarrow \quad \frac{9c^2-2ct+57t^2}{64t}\geq\frac{t}{100}\left[\frac{337}{2}-36\frac{c}{t}+19\left(\frac{c}{t}\right)^2\right]+\frac{t}{8}-2\frac{(5t-c)^2}{32t} \\ &\Leftrightarrow \quad \left(\frac{149}{64}-\frac{337}{200}\right)t+\left(\frac{9}{25}-\frac{21}{32}\right)c+\left(\frac{13}{64}-\frac{19}{100}\right)\frac{c^2}{t}\geq 0.\end{aligned}$$

Note first that the last bracket $\left(\frac{13}{64}-\frac{19}{100}\right)$ is positive. Moreover, since $c\leq t$, the LHS is larger than

$$\left(\frac{149}{64}-\frac{337}{200}\right)c+\left(\frac{9}{25}-\frac{21}{32}\right)c=\left(\frac{107}{64}-\frac{265}{200}\right)c>0 \tag{6}$$

The LHS is therefore positive so the DS's revenue is higher in the benchmark case where only one firm can benefit from some information about consumers.

Proof of Lemma 5

In equilibrium we must have $p_A = p_A^a$ and $p_B = p_B^a$, in which case $\theta_2 < \theta_1$. However, reaction functions should be calculated by taking consumer actions as given. In case (i) B's reaction function is given by

$$\arg\max_{p_B} m[1-\theta_1(p_A,p_B)]p_B,$$

and A's reaction function is

$$\arg\max_{p_A} [m\theta_1(p_A,p_B)+\theta_2(p_A^a,p_B^a)]p_A.$$

These two functions allow us to find $p_A(\theta_2)$ and $p_B(\theta_2)$. Finally, because $p_A = p_A^a$ and $p_B = p_B^a$ in equilibrium, we can plug these reaction functions into equation (4); doing so leads to the expression for θ_2.

For case (ii) we solve in the same manner but change B's reaction function to account for selling also to old consumers:

$$\arg\max_{p_B} [(1-\theta_2^0(p_B))+m(1-\theta_1(p_A,p_B))]p_B.$$

Recall that θ_2^0 (i.e., B's market share among old consumers) depends on p_B but not on p_B^a; the reason is that θ_2^0 is set by $p_A(\theta_2^0)=0$. And since $p_A(\theta)=p_B+(1-2\theta)t$ is always chosen by A at stage 4, it follows that θ_2^0 is a function of p_B and not of p_B^a. Therefore, B optimizes in both markets at the same time. We must assume that $c \leq \frac{t(3m+3)}{3m+4}$ in order to ensure that a positive proportion of consumers pays the privacy cost.

Proof of Lemma 6

Compute the difference in B's profits when B does not sell to old consumers (case (i) in Lemma 3) and when B does sell to old consumers (case (ii) in Lemma 3):

$$\Delta\pi = \frac{m(c-(3m+2)t)^2}{2(3m+1)^2 t} - \frac{\left(-cm+3m^2t+4mt+t\right)^2}{2(m+1)(3m+2)^2 t}.$$

The first term on the RHS represents B's profits when it does not sell to old consumers; the second term is B's profits when it does. It can be verified that that $\Delta\pi \to t/6$ as $m\to\infty$. Moreover, $\Delta\pi \to -t/8$ as $m \to 0$. It can also be proved that $\Delta\pi$ is always increasing. Therefore, a threshold $m^* > 0$ exists.

Proof of Proposition 9

If $m \geq m^*(c)$, then B does not compete for old consumers. So just as in the proof of Proposition 6, it suffices to show that $\pi_A - \pi_B \geq 2\big(\frac{t(2m+1)}{4} - \pi_B\big)$ for

$$\pi_A = m \int_0^{\frac{p_B - p_A + t}{2t}} p_A \, d\theta + \int_0^{\theta_2} p_A \, d\theta + \int_{\theta_2}^1 (p_B + (1-2\theta)t) \, d\theta$$

and

$$\pi_B = m \int_{\frac{p_B - p_A + t}{2t}}^1 p_B \, d\theta,$$

where p_A, p_B, and θ_2 are as given in Lemma 3 (i).

If $m < m^*(c)$, then B does supply old consumers. Again as in the proof of Proposition 6, it suffices to show that $\pi_A - \pi_B \geq 2\big(\frac{t(2m+1)}{4} - \pi_B\big)$ for

$$\pi_A = m \int_0^{\frac{p_B - p_A + t}{2t}} p_A \, d\theta + \int_0^{\theta_2} p_A \, d\theta + \int_{\theta_2}^{\frac{p_B + t}{2t}} (p_B + (1-2\theta)t) \, d\theta$$

and

$$\pi_B = m \int_{\frac{p_B - p_A + t}{2t}}^1 p_B \, d\theta + \int_{\frac{p_B + t}{2t}}^1 p_B \, d\theta,$$

where p_A, p_B, and θ_2 are as given in Lemma 3 (ii).

Appendix C: Imperfect tracking technology

Here we consider the case of an imperfect tracking technology. More precisely, following Belleflamme (2016), we assume that absent any protection, a consumer's characteristics will be known by the data broker with probability $\lambda \in [0, 1]$. We will show that our main results—in particular, the exclusivity result—are preserved under this modification. We focus here on the analysis of the duopoly case.

First we consider the case where the two firms have acquired the information and show that the result of Lemma 1—which states that the privacy cost is never paid by consumers—still holds. We know that for consumers whose information has been acquired by both firms, the prices proposed by firms A and B will be as given in Proposition 1. But now, these prices are relevant to only a share λ of the consumers who have not paid the privacy cost. The other share $1 - \lambda$ can, from the firm's perspective, be pooled with new consumers. Retaining the notation of Lemma 1, old consumers who have paid the privacy cost are of types in $[0, \theta_A]$ or $[\theta_B, 1]$.

When choosing its basic price, A solves

$$\max_{p_A} \lambda \int_0^{\theta_A} p_A \, d\theta + (2-\lambda) \int_0^{\frac{1}{2}+\frac{p_B-p_A}{2t}} p_A \, d\theta.$$

Firm B solves an analogous problem, after which the prices are given by $p_A = t + \frac{2t}{3}\frac{(2\theta_A+1-\theta_B)}{2-\lambda}$ and $p_B = t + \frac{2t}{3}\frac{(\theta_A+2(1-\theta_B))}{2-\lambda}$.

The condition under which consumers pay for privacy differs slightly from the one used in the main model, because not paying now leads to type revelation only with probability λ. For consumers to the left of the Hotelling line, this modified condition is written as $v - p_A - \theta t - c \geq \lambda[v - p_A(\theta) - \theta t] + (1-\lambda)[v - p_A - \theta t]$, or equivalently, as $\theta \leq \theta_2 = \frac{t - p_A - \frac{c}{\lambda}}{2t}$ (where we have used $p_A(\theta) = (1-2\theta)t$).

Because $p_A = t + \frac{2t}{3}\frac{(2\theta_A+1-\theta_B)}{2-\lambda} > t$, no consumers with positive c wants to pay the privacy cost in this case. Therefore, a firm buying the data from the broker will be fully informed about the share λ of old consumers. For the other group of consumers (of mass $(2-\lambda)$ and uniformly distributed), the firms will compete à la Hotelling. The two firms' profit will then be $\pi_A = \pi_B = (2-\lambda)t - \lambda\frac{t}{4}$. Let $M = (2-\lambda)/\lambda$. Then, the profit is $\lambda[M\frac{t}{2} + \frac{t}{4}]$, which is proportional to the profit found in Section 5.2 where markets have different sizes. The imperfection of the tracking technology tends to increase the mass of consumers for which no information is available, an effect that is similar to an increase in the mass of new consumers.

We now consider the case where only firm A has acquired the information. Firm B may or may not have decided to compete for the consumers about which firm A has some information about; however, that decision does not affect how A chooses its basic price. It is the solution of the program $\max_{p_A}[(2-\lambda)\theta_1(p_A, p_B) + \lambda\theta_2]p_A$ or equivalently $\max_{p_A} \lambda[M\theta_1(p_A, p_B) + \theta_2]p_A$, where $M = \frac{(2-\lambda)}{\lambda}$. This leads to a reaction function given by $p_A = \frac{t+p_B}{2} + \frac{t\theta_2}{M}$. In this case, the personalized price that A chooses is given by $p_A(\theta) = p_B + (1-2\theta)t$ (when that value is positive). The condition under which consumers pay the privacy cost is therefore $\theta \leq \theta_2 = \frac{t+p_B-p_A-\frac{c}{\lambda}}{2t}$.

If firm B does not compete for the consumers that firm A has some information about, then its basic price can attract new consumers and a share $1-\lambda$ of old consumers. Firm B's objective is $\max_{p_B}(2-\lambda)(1-\theta_1(p_A,p_B))p_B$ and the reaction function is $p_B = \frac{p_A+t}{2}$. Continuing to put $M = \frac{(2-\lambda)}{\lambda}$, we see that the equilibrium basic prices are given by the expressions in Lemma 3(i) but with c/λ instead of c and with M instead of m.

Now suppose that firm B does try to compete for consumers about which firm A has some information. In this case, a portion of the share λ of old consumers who have not paid the privacy cost will be served by firm B. These consumers are of type $\theta \in [\theta_2^0, 1]$ where $\theta_2^0 = \frac{p_B+t}{2t}$. So firm B's objective writes as

$$\max_{p_B}(2-\lambda)(1-\theta_1(p_A,p_B))p_B + \lambda(1-\theta_2^0)p_B$$

and the reaction function is $p_B = \frac{p_A + t(1+M)}{4}$. Here the equilibrium basic prices are given by the expressions in Lemma 3(ii) but with c/λ and M replacing (respectively) c and m.

This extension with an imperfect tracking technology is therefore similar to the one studied in Section 5.2. A change in the value of λ has two effects. First it changes the value of the privacy cost—although this change is the same for all expressions of the profit. Second, it modifies the relative size of the two markets exactly as described in Section 5.2. This means that, for a given value of λ (hence of c), the ranking of the profits derived in Section 5.2 is unaffected. Proposition 8's exclusivity result, whereby the broker sells consumer data to just one firm no matter what the values of c and m, is also valid for any value of λ.

Additional Reference

Belleflamme, P. (2016). "Monopoly Price Discrimination and Privacy: The Hidden Cost of Hiding", *Economics Letters*, Vol. 149, 141-144.