Data Purchase and Third Degree Price Discrimination in Duopoly

Preliminary Version

Christos Constantatos[1] and Apostolos Ioannis Martis[1]

[1]University of Macedonia

cconst@uom.edu.gr
amartis@uom.edu.gr

April 2025

Abstract

This paper analyzes strategic data purchasing decisions by firms in a differentiated duopoly market where individual consumers have a continuous demand (no discrete-choice decisions). In contrast to previous literature, data acquisition from both firms is the unique equilibrium of the game. The welfare analysis demonstrates that symmetric data acquisition maximizes both individual and total industry profits, but reduces consumer surplus and total welfare. These findings provide critical insights for competition authorities considering data market regulations and highlight the tension between firm profitability and social welfare in data-driven markets.

Keywords: Price Discrimination; Representative Consumer; Duopoly; Consumer Data
JEL Codes: D21; L11; L13

1 Introduction

As digital markets increasingly rely on consumer data for competitive advantage, understanding firms' strategic incentives for data acquisition becomes crucial for competition policy and market regulation. Mikians et al. (2013)acknowledged the presence of price variation, ranging from 10% to 30%, for certain retailers offering the same product. This disparity in pricing suggests that some retailers may be engaging in discriminatory practices, charging higher prices to specific groups of customers.. Hannak et al. (2014) collected user data from Amazon's Mechanical Turk (AMT) to examine for price discrimination. Their analysis revealed evidence of price discrimination in various sectors, including general retailers and traveling agencies. This phenomenon raises important questions about the fairness and transparency of pricing practices, particularly in online marketplaces where consumers may be more vulnerable to discriminatory pricing. Finally, Woollaston (2014) present real-life price discrimination cases, based on consumer data. For example, in 2010, Amazon was found to be selling identical products at varying prices. This pricing disparity was also observed in 2012 by Staples, which employed a pricing model that differentiated costs among consumers based on their geographical location.

In this paper, we develop a differentiated on-line duopoly model in order to examine the effects of data acquisition on prices, profits and welfare. On the demand side, each individual consumer's quantity demanded varies continuously with prices. For instance, one may think of the good being cosmetics, cloths, or other accessories, rather than computers or mobile phones.[1] We use a representative-consumer model, assuming, however, a continuum of representative-consumer types, distributed according to their willingness-to-pay (WTP) for the product. One may think that there is a large type of sub-markets, each represented by consumers with a specific utility function. The latter is associated to some particular socioeconomic characteristics of each consumer that are initially unknown to the two firms.

On the supply side, two downstream firms engage in price competition. An independent data broker collects consumer data from various sources and offers them to downstream firms at a fixed price. Since sales are on-line, access to those data enables a firm to engage in third-degree price discrimination by setting personalized prices, based on each representative consumers' valuation for the products (we assume that product reselling among consumers is very costly.) In the baseline model, we assume that consumers are passive, in the sense that they do not pay for privacy, i.e., no possibility to be excluded from the data set even at a cost; later on we extend the model by allowing such possibility. We analyze three distinct regimes: (i) Bertrand Competition, in which both firms charge uniform prices; (ii) Data Exclusivity, where only one firm purchases consumer data; and (iii) Data Acquisition, where both firms obtain access to consumer data.

Our model reveals that in the downstream market both firms wish to buy consumer data and the supplier's optimal strategy is to sell the dataset to both parties. This finding contrasts with the results in Montes, Sand-Zantman, and Valletti (2019), where it is shown that the upstream firm typically engages in exclusive data sales to only a single downstream firm. The reason for this difference lies in the fact that they consider a Hotelling model where each consumer makes a discrete decision about purchasing one unit or nothing. Furthermore, our model shows that while aggregate quantities remain invariant across regimes, the average price may vary. When both firms acquire data the average price increases, as expected. What is less expected is that in the Exclusivity scenario, the average price of the firm that purchases the data remains the same as in the no-data Bertrand equilibrium: exclusive data purchase only allows the data-possessing firm to offer a price spread adapted to individual WTP. This implies that in the case of Exclusivity, there are two effects on the profit of the data possessing firm: a positive one, related to the practicing of third-degree price discrimination, but also a ***negative one*** due to the fact that in the high WTP markets its product is more expensive compared to that of its rival who only sets a single price for all the markets.[2] The negative effect disappears when the rival also acquires data and becomes able to price-discriminate. Thus, we arrive to the surprising conclusion that, while overall, exclusive data acquisition by one firm increases its profits, when the rival firm also acquires data the profit of the first acquirer instead of being lowered, it *further increases!* In conclusion, while compared to the no-data Bertrand equilibrium, exclusive acquisition by a single firm increases the profits of that firm while leaving those of the non-acquirer unchanged, when both firms acquire data individual profits are even higher. The story is of course quite different for consumers and

[1]In the conclusions we explain how the model can be interpreted to include such discrete-choice goods.

[2]Of course, in the low valuation markets the firm that owns data charges a lower price, yet these markets yield less profit .

total surplus. Exclusive acquisition reduces total surplus because it induces price discrimination without increase in quantities (see Tirole, 1988). Acquisition by both firms makes things even worse because it also results in higher prices.

There is a substantial body of literature on data intermediaries, consumer data, and price discrimination (see Acquisti, Taylor, and Wagman (2016); Pino (2022) for a comprehensive literature review). Consumer data create a trade-off for consumers between improved product matching and the risk of price discrimination. Ichihashi (2020) show that a firm may credibly commit to using consumer data solely for recommendations rather than for personalized pricing. As is well known, downstream firms primarily use consumer data for price discrimination. Acquisti and Varian (2005) discuss behavior-based price discrimination, where firms condition prices on consumers' past purchasing behavior, and argue that this may improve consumer welfare compared to uniform pricing. However, as Taylor and Wagman (2014) emphasize, the welfare implications of data usage depend on the specific modeling framework. Bounie, Dubus, and Waelbroeck (2021) examine the incentives of a data broker to sell data exclusively or share it across firms and find that only high-WTP consumers' data are sold, and typically to a single firm. Closest to our work, Montes, Sand-Zantman, and Valletti (2019) use a discrete-choice model to study downstream competition under alternative data-sharing regimes and show that consumer surplus is maximized when both firms access the data, whereas industry profits are maximized when only one firm does—though this leads to the lowest total welfare. A key difference between our analysis and theirs is the demand structure: they model discrete-choice goods (e.g., computers), where all consumers have the same WTP for one unit and data reveal proximity to a firm in a Hotelling framework, enabling full surplus extraction absent competition. In contrast, we focus on goods with continuous demand (e.g., clothes, cosmetics), where data—such as demographics, education, and past purchases—reveal information about consumers' underlying WTP.

The structure of the paper is as follows: section 2 introduces the baseline model, section 3 presents the three scenarios examined and derives the equilibrium. Finally, section 4 concludes.

2 The Model

We consider a downstream market with two symmetric single-product firms selling differentiated products. Each firm faces constant marginal production cost which is assumed equal to zero, for simplicity. There is also an upstream firm who obtains a dataset and it is able to sell it to one or both firms[3]. Individual consumers purchase the two goods at varying quantities, depending upon the two products' prices.[4] The utility function of consumer i is

$$U_i = k_i \alpha (q_{1i} + q_{2i}) - \frac{b}{2}(q_{1i}^2 + 2\theta q_{1i} q_{2i} + q_{2i}^2) \tag{1}$$

where $k_i, \alpha, b > 0$, $0 \leq \theta \leq 1$, and q_{ji}, $j = 1, 2$, represents the quantity of good j purchased by individual i. The parameter k_i is idiosyncratic, representing the intensity of each individual consumer's willingness-to-pay. We assume that consumers are distributed uniformly according to $k \in [\underline{k}, 1]$. The exact value of k_i is known to consumer i but initially unknown

[3] We assume that the cost of data acquisition for the upstream firm is 0

[4] Even at the individual level, the purchase decision is not "one unit or nothing," but depends smoothly upon the product's price. Examples are clothes, accessories, cosmetics, detergents, etc., rather than computers or cars.

to the firms. However, by paying a cost T to a third party, each firm can purchase a data set that enables it to identify consumers according to k.

Maximizing Equation 1 leads to the demand functions for the two goods:

$$D_{ji}(p_{ji},\ p_{-ji},\ k_i) = \frac{k_i\alpha(1-\theta) - p_{ji}(k_i) + \theta\, p_{-ji}(k_i)}{b(1-\theta^2)} \tag{2}$$

Since in the absence of data, a firm cannot price discriminate, that firm must maximize profit considering the sales that corresponding to its aggregate demand at any single posted price. Thus, its profit function is given by:

$$\Pi_j^N = p_j \int_{k_i\in[\underline{k},1]} D_{ji}(p_j,\ p_{-ji},\ k_i)\ f(k_i)\ dk_i \tag{3}$$

where p_{-ji} is rival price at the market characterized by k_i and where $f(k_i) = \dfrac{1}{1-\underline{k}}$ is the density function.

We exclude the possibility of reselling among consumers, therefore data acquisition allows the acquiring firm to practice 3rd-degree price discrimination by posting a different price at each k_i market. If firm j can charge personalized prices, p_{ji}, its reaction function depend on the uniform price set by the other firm. This price is calculated based on the average consumer. On the contrary, when both firms can discriminate, according to consumer i's willingness to pay (WTP), each firm's reaction function depends on the price imposed by the other firm for a specific k_i. Therefore, firms engage in a Bertrand competition not for the average consumer, but for each type. In this case, firms' profits are:

$$\Pi_j^D = \int_{k_i\in[\underline{k},1]} p_{ji}\ D_{ji}(p_{ji},\ p_{-ji},\ k_i)\ f(k_i)\ dk_i \tag{4}$$

We respect to downstream equilibrium, we consider three scenarios: Bertrand Competition (no data acquisition), Data Exclusivity, and Data Acquisition (denoted as *NN*, *DN* and *DD* , respectively). In the NN equilibrium, both firms maximize (3), whereas in the DD equilibrium, both firms maximize (4); in the asymmetric DN cases (two symmetric equilibria) the data possessing firm maximizes (4) and the one that has no data maximizes (3). When firms are asymmetric, since the firm that possesses no data charges a single price for all the markets, its optimal response must take into account that in equilibrium that firm may not sell in all the markets.[5] In this part of the work, we rule out this possibility by assuming that $\underline{k}$ is sufficiently high for no market in the DN equilibrium to yield negative quantity for the firm that has no data.
Notice, that when firms obtain data and when they don't, their profits are multiplied with the density function. This is because, $f(k_i)\ dk_i$ represents the proportion of total consumers who are of type k_i. So, when aggregating firm 1's total profits over the entire population one need to weigh each type by how frequently it occurs, which is what $f(k_i)$ does. In what follows, we first consider that consumers cannot pay a privacy cost

3 No privacy cost

In this section, we present the equilibrium absent of consumers' privacy decision in each of the scenarios presented in section 2, using backward induction, and then compare them regarding

[5] In some low k markets, the universal price that firm charges may yield negative sales.

total industry profits, consumer surplus, and total welfare. The timing of the game is as follows: First, the upstream firm offers access to the database at a fixed cost (to one or both firms). At the second stage, firms decide whether or not to purchase the data. Finally, firms set their prices.

3.1 Scenario 1: Bertrand Competition – *NN*

Since firms are symmetric, the Bertrand case concerns maximization of Equation 3 with respect to $p_j,\ j = 1, 2$.

$$\underset{p_j \geq 0}{argmax} \left\{ \Pi_j = \ p_j \int_{\underline{k}}^{1} \frac{k_i\ \alpha(1-\theta) - p_j + \theta\ p_{-j}}{b(1-\theta^2)} f(k_i)\ dk_i \right\}$$

The optimal prices and quantities are:

$$p_j^{NN} = \frac{\alpha(1-\theta)(1+\underline{k})}{2(2-\theta)},\ \ q_i^{NN} = \frac{\alpha[2(2-\theta)k_i - (1-\theta)(\underline{k}+1)]}{2b(2-\theta)(\theta+1)},\ \ Q^{NN} = \frac{\alpha\,(\underline{k}+1)}{2b(2-\theta)(\theta+1)}$$

with $\underline{k} \geq \dfrac{1-\theta}{3-\theta}$. The latter arises when considering $q^*_{i,NN} \geq 0$. First of all, note that as products become close substitutes, price monotonically decreases, as it should. However, the trajectory of total quantity does not exhibit a monotonic trend: for $\theta \leq (\geq)\dfrac{1}{2}$, an increase in θ lowers (increases) quantity. This phenomenon can be attributed to the fact that when products are not highly substitutable, as θ increases, the interval in which $\underline{\text{k}}$ belongs to decreases in width, ultimately leading to a reduction in active consumers in equilibrium. Conversely, when θ exceeds $1/2$, small increases in substitutability stimulate an increase in quantity as firms become more aggressive in their pricing strategies. By substituting the aforementioned quantities and prices to Equation 3, we obtain the firms' and the total industrial profits. Moreover, since demands are linear in prices, it is easy to compute the consumer surplus for a given price of the other firm. Adding the industrial profits and the consumer surplus leads to total welfare. All of them are presented in the following lemma.

Lemma 1. *For all $\theta \in [0, 1]$ and $\underline{k} \geq \dfrac{1-\theta}{3-\theta}$, the firms' and the industry's profits, the consumer surplus and total welfare in the Bertrand Competition are:*

$$\Pi_j^{NN} = \frac{\alpha^2(1-\theta)(\underline{k}+1)^2}{4b(2-\theta)^2(\theta+1)},\ j = 1, 2$$

$$T\Pi_{NN} = \frac{\alpha^2(1-\theta)(\underline{k}+1)^2}{2b(2-\theta)^2(\theta+1)}$$

$$TCS_{NN} = \frac{\alpha^2\left[(\theta^2-4\theta)\left(1-\underline{k}\right)^2 + 7\underline{k}^2 - 2\underline{k} + 7\right]}{12b(2-\theta)^2(\theta+1)}$$

$$TW_{NN} = \frac{\alpha^2\left\{[(\theta-10)\theta+13]\underline{k}^2 - 2[\theta(\theta+2)-5]\underline{k} + (\theta-10)\theta + 13\right\}}{12b(2-\theta)^2(\theta+1)}$$

Proof. Firms' and industry profits in equilibrium can be easily calculated by using $\Pi_j^{NN} = p^*_{NN}\ Q^*_{NN} = \dfrac{\alpha(1-\theta)(1+\underline{k})}{2(2-\theta)}\ \dfrac{\alpha\,(\underline{k}+1)}{2b(2-\theta)(\theta+1)}$ and then $T\Pi_{NN} = 2\Pi_j^{NN}$. To calculate the

TW, since firms do not carry out any costs, we substitute $q^*_{i,\,NN}$ to Equation 1 and integrate to the active consumer segment, i.e., from $\underline{k}$ to 1. Finally, total consumer surplus is merely $TCS_{NN} - T\Pi_{NN}$. $\square$

Notably, a negative correlation is observed between total profits and θ, as the increased substitutability of products necessitates more intense price competition among firms. In contrast, consumer surplus (CS) exhibits a U-shaped relationship with θ. Specifically, at low levels of substitutability, firms operate in a market characterized by limited price competition, resulting in elevated prices and a subsequent reduction in quantity demanded. Conversely, at high levels of θ, products become immediate substitutes, leading to increased price competition and a corresponding increase in CS. Furthermore, the total welfare is observed to decline sharply at low levels of θ, as the reduction in profits outweighs the growth in CS. However, as θ increases, CS eventually surpasses total profits, resulting in a more gradual decline in total welfare (for a graphic illustration of CS, $T\Pi$ and TW, see Figure 5).

3.2 Scenario 2: Data Exclusivity – *DN*

Now, suppose that the upstream firm offers an exclusive access only to one firm. This scenario may arise when a supplier firm maintains an exclusive relationship with a single downstream entity, or when ethical considerations preclude data sharing. Consequently, the firm in possession of the data can leverage this information to offer tailored pricing, whereas its competitor is relegated to charging a uniform price. Here, we assume, without loss of generality, that Firm 1 obtains consumer data. Therefore, it is capable of providing tailored prices to consumers, whereas Firm 2 continues to offer uniform pricing schemes. Formally, the two firms maximize their profits with respect to p_{1i} for firm 1 and p_2 for firm 2. Mathematically, firms 1 and 2 face the following problem:

$$\underset{p_{1i}\geq 0}{max}\left\{\Pi_1 = \int_{\underline{k}}^{1} p_{1i}\,\frac{\alpha k_i(1-\theta) - p_{1i}\,(k_i) + \theta\, p_2}{b\,(1-\theta^2)}\,\frac{1}{1-\underline{k}}\,dk_i\right\}$$

and

$$\underset{p_{2}\geq 0}{max}\left\{\Pi_2 = p_2\int_{\underline{k}}^{1} \frac{\alpha k_i(1-\theta) - p_2 + \theta\, p_{1i}\,(k_i)}{b\,(1-\theta^2)}\,\frac{1}{1-\underline{k}}\,dk_i\right\}$$

Before proceeding to the solution of the subgame, it is worth presenting firm 2's reaction function. The first order condition gives:

$$p_2^{RF}\,(p_{1i}) = \frac{1}{4}\alpha(1-\theta)\,(\underline{k}+1) + \frac{\theta\displaystyle\int_{\underline{k}}^{1} p_{1i}\,(k_i)\;dk_i}{2(1-\underline{k})}$$

It is essential to note that Firm 2 responds to the average of Firm 1's personalized prices within the segment $[\underline{k}, 1]$. Although the optimal response does not involve tracking each individual price, any alteration to a specific price $p_{1i}(k_i)$ results in a shift in the average, consequently influencing Firm 2's reaction. For example, mild competition (thus increased prices) tend to increase firm 2's price, for a given θ. For $\underline{k} \geq \dfrac{2-\theta^2}{6-\theta^2}$, the prices and quantities (individual and aggregate) are:

$$p_{1i}^{DN}(k_i) = \frac{1}{4}\alpha(1-\theta)\left[\frac{\theta(\underline{k}+1)}{2-\theta} + 2k_i\right],\; p_2^{DN} = \frac{\alpha(1-\theta)\,(\underline{k}+1)}{2(2-\theta)}$$

$$q_{1i}^{DN} = \frac{\alpha\left[\theta\left(\underline{k}-2k_i+1\right)+4k_i\right]}{4b(2-\theta)(\theta+1)}, \; q_{2i}^{DN} = \frac{\alpha\left[2\left(4-\theta^2\right)k_i-\left(2-\theta^2\right)\left(\underline{k}+1\right)\right]}{4b(2-\theta)(\theta+1)}$$

$$Q_1^{DN} = Q_2^{DN} = \frac{\alpha\left(\underline{k}+1\right)}{2b(2-\theta)(\theta+1)}$$

There are several notable aspects worth highlighting in this context. First, firm 2 sets the same price as in *NN* case. This is due to Firm 2 facing an average demand, and its optimal response is contingent solely on the aggregate distribution of Firm 1's pricing schedule. Since Firm 1's average price across types matches the *NN* equilibrium, Firm 2 does not change the price [6]. Figure 1 shows that firm 2's price is equal to the average Firm 1's price across all types of consumers. Due to the aforementioned result, aggregate quantities remain the same as well.

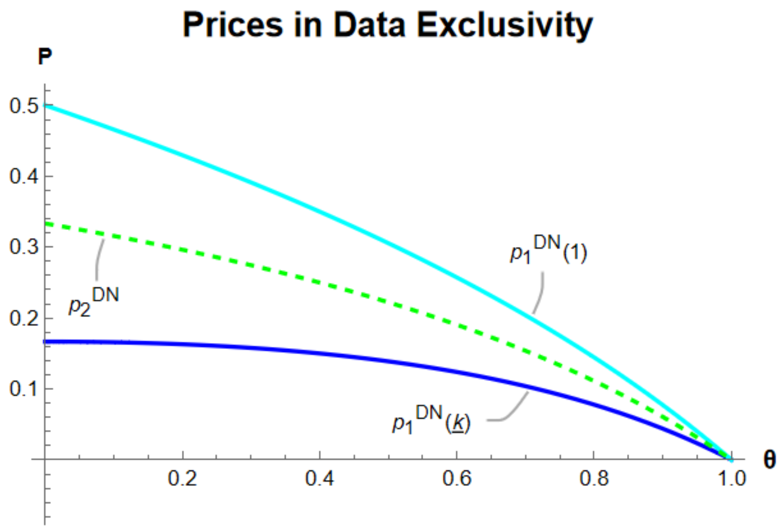


Figure 1: Prices in Data Exclusivity

The equilibrium profits, consumer surplus and welfare are presented in the following lemma:

Lemma 2. *For all $\theta \in [0,1]$ and $\underline{k} \geq \frac{2-\theta^2}{6-\theta^2}$, the firms' and the industry's profits, the consumer surplus and total welfare in the Bertrand Competition are:*

$$\Pi_1^{DN} = \frac{\alpha^2(1-\theta)\left[\left(\theta^2-4\theta\right)\left(1-\underline{k}\right)^2+16\left(\underline{k}^2+\underline{k}+1\right)\right]}{48b(2-\theta)^2(\theta+1)}, \; \Pi_2^{DN} = \frac{\alpha^2(1-\theta)\left(\underline{k}+1\right)^2}{4b(2-\theta)^2(\theta+1)}$$

$$T\Pi_{DN} = \frac{\alpha^2(1-\theta)\left[\left(\theta^2-4\theta\right)\left(1-\underline{k}\right)^2+4\underline{k}\left(7\underline{k}+10\right)+28\right]}{48b(2-\theta)^2(\theta+1)}$$

$$TCS_{DN} = \frac{\alpha^2\left[\left(3\theta^3-7\theta^2-8\theta\right)\left(1-\underline{k}\right)^2+44\underline{k}^2+8\underline{k}+44\right]}{96b(2-\theta)^2(\theta+1)}$$

$$TW_{DN} = \frac{\alpha^2\left\{\left(\theta^3+3\theta^2\right)\left(\underline{k}-1\right)^2-24\theta\left[\underline{k}\left(3\underline{k}+2\right)+3\right]+4\left[\underline{k}\left(25\underline{k}+22\right)+25\right]\right\}}{96b(2-\theta)^2(\theta+1)}$$

[6] This result may be due to linear demand functions and uniform distribution of k_i

Proof. Π_1^{DN} and Π_2^{DN} are calculated by multiplying their respective prices with the aggregate quantity, i.e., $\Pi_1^{DN} = \int_{\underline{k}}^{1} \frac{1}{4}\alpha(1-\theta)\left[\frac{\theta(\underline{k}+1)}{2-\theta}+2k_i\right]\frac{\alpha\left[\theta\left(\underline{k}-2k_i+1\right)+4k_i\right]}{4b(2-\theta)(\theta+1)}\frac{1}{1-\underline{k}}dk_i$ and $\Pi_2^{DN} = \frac{\alpha(1-\theta)\left(\underline{k}+1\right)}{2(2-\theta)}\frac{\alpha\left(\underline{k}+1\right)}{2b(2-\theta)(\theta+1)}$. Now, since there are no costs on the production side, welfare is calculated by substituting q_{1i}^{DN} and q_{2i}^{DN} on Equation 1 for $\underline{k} \geq \frac{2-\theta^2}{6-\theta^2}$ which ensures positive individual quantities. When we substitute the total profits from the welfare function, consumer surplus arises. □

Since prices and aggregate quantities for firm 2 are the same as in Bertrand Competition, its profits remain the same. Now, even though individuals consume same quantities from both of the firms, it is shown that since firm 1 can set individualized prices makes it have higher profits than that of firm 2 (see Appendix A). Furthermore, as θ increases, profits exhibit a tendency towards equivalence with marginal cost (which here is zero). Like in the case of *NN*, welfare and total profits monotonically decrease. Consumer surplus maintains a non-monotonic behavior, with it being positively-sloped for low θ and negatively-sloped otherwise.

3.3 Scenario 3: Data Acquisition – *DD*

In this scenario, both companies have access to consumer data, which enables them to fully distinguish consumers. Essentially, in this case, they compete for every consumer in the $[\underline{k}, 1]$ interval. Therefore, firms engage in a Bertrand Competition game for each type of consumer. Mathematically, both firms will maximize profits as described in Equation 4:

$$\max_{p_{ji}\geq 0} \Pi_j = \int_{\underline{k}}^{1} p_{ji}\, D_{ji}(p_{ji},\, p_{-ji},\, k_i)\, dk_i,\ j = 1, 2$$

The prices and quantities that arise in equilibrium for both firms are:

$$p_{ji}^{DD} = \frac{\alpha(1-\theta)k_i}{2-\theta},\ q_i^{DD} = \frac{\alpha k_i}{b(2-\theta)(\theta+1)},\ Q_j^{DD} = \frac{\alpha\left(\underline{k}+1\right)}{2b(2-\theta)(\theta+1)}$$

Once again, notice that the aggregate quantity is the same as both *NN* and *DN*. This is a result of linear demand functions and uniform distributions. However, the prices and individual quantities are different. The profits, consumer surplus and total welfare in equilibrium are:

Lemma 3. *For all $\theta \in [0, 1]$, the firms' and the industry's profits, the consumer surplus and total welfare in the Bertrand Competition are:*

$$\Pi_j^{DD} = \frac{\alpha^2(1-\theta)\left(\underline{k}^2+\underline{k}+1\right)}{3b(2-\theta)^2(\theta+1)}$$

$$T\Pi_{DD} = \frac{2\alpha^2(1-\theta)\left(\underline{k}^2+\underline{k}+1\right)}{3b(2-\theta)^2(\theta+1)}$$

$$TCS_{DD} = \frac{\alpha^2\left(\underline{k}^2+\underline{k}+1\right)}{3b(2-\theta)^2(\theta+1)}$$

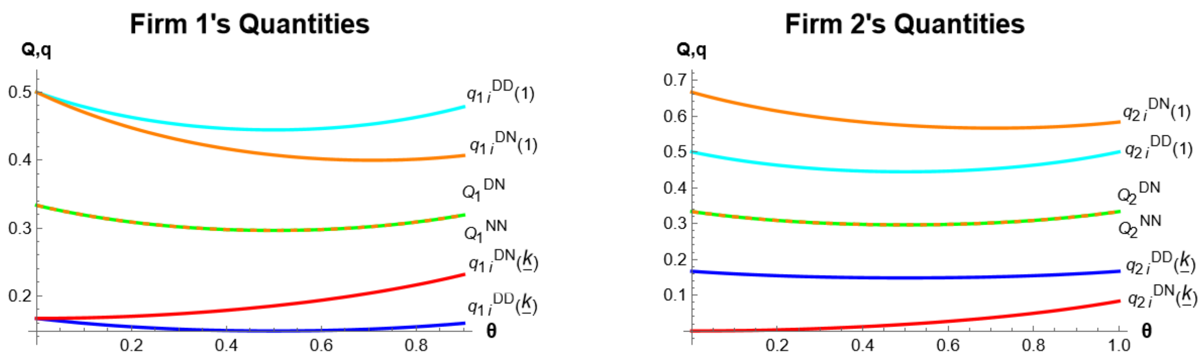
$$TW_{DD} = \frac{\alpha^2(3-2\theta)\left(\underline{k}^2+\underline{k}+1\right)}{3b(2-\theta)^2(\theta+1)}$$

Proof. Each firms' profits can be calculated as $\Pi_j = \displaystyle\int_{\underline{k}}^{1} \frac{\alpha(1-\theta)k_i}{2-\theta}\frac{\alpha k_i}{b(2-\theta)(\theta+1)}\frac{1}{1-\underline{k}}\,dk_i$. Total profits is merely the addition of the individual profits. For TW_{DD}, we substitute q_i^{DD} to Equation 1, since the firms do not have any costs. Finally, consumer surplus is calculates as $TW_{DD} - T\Pi_{DD}$. □

The behavior of the aforementioned functions exhibits a consistent pattern across the examined scenarios. A negative correlation exists between industry profits and θ, whereas consumer surplus initially decreases before experiencing an increase at a certain threshold of θ. Consequently, the total welfare metric, which combines total profits and consumer surplus, displays a sharp decline for initial levels of θ due to the concurrent decrease in total profits and, to a lesser extent, consumer surplus. However, as consumer surplus surpasses industry profits, the rate of decline in total welfare moderates. In conclusion, total welfare and profits demonstrate a decline relative to θ, whereas consumer surplus exhibits a non-monotonic trajectory. The subsequent subsection will focus on a comparative analysis of the three scenarios, with the aim of identifying the equilibrium in the downstream market.

3.4 Comparisons

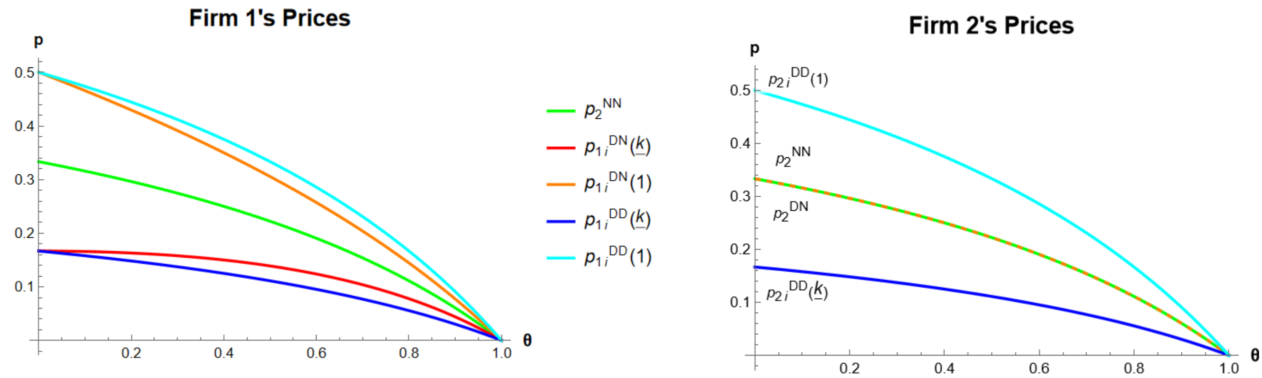
A comparative analysis of equilibrium variables is now feasible, allowing for a nuanced understanding of their relative significance. The forthcoming figures employ the following color-coding scheme: blue and cyan for DD, orange and red (dashed and solid lines) for DN, and green (dashed and solid lines) for NN. Initially, we will compare elementary variables, specifically prices and quantities. Notably, aggregate quantities for both firms 1 and 2 are identical, with $Q_j^{NN} = Q_j^{DN} = Q_j^{DD} = \dfrac{\alpha(\underline{k}+1)}{2b(2-\theta)(\theta+1)}$. The sole distinction lies in the individual quantities supplied by firm 1 under the Data Exclusivity scenario. As depicted in Figure 2, quantities in the Data Exclusivity scenario surpass those of the Data Acquisition scenario for consumers with low willingness-to-pay, whereas they are lower for high-valuation consumers. Conversely, firm 2 sells higher quantities to high-WTP consumers and lower quantities to those with lower WTP. In conclusion, when a firm can price discriminate while the other cannot, the former "wins" over low-valuation consumers while the other does so on high-WTP ones.



(a) Firm 1 Quantity Comparison (b) Firm 2 Quantity Comparison

Figure 2: Firms 1 and 2 Quantity Comparison – Individual and Aggregate Quantities

Those results are contingent to the prices in equilibrium. As depicted in Figure 3a firm 1 charges a low price for high valuation consumers in *DN* than in *DD* case, whereas prices are higher for the low valuation consumers. However, from the previous section, it was shown that p_2^{DN} is equal to p_j^{NN} which is the average price charged by the firms when none of them has access to consumer data, meaning that in the DN case, firm 2 gains high-valuation. When we are on the DD case, the price spectrum increases. (from red-orange range to blue-cyan, for firm 1, and from orange-dashed line to blue-cyan range for firm 2). Essentially, this leads to higher prices for high-WTP consumers and lower for low-WTP ones. Since, however, both firms want to attract the former, the high prices ultimately lead to higher prices.



(a) Firm 1 Price Comparison (b) Firm 2 Price Comparison

Figure 3: Firms 1 and 2 Price Comparison – Personalized and Uniform Prices

The following proposition shows the equilibrium in the downstream market.

Proposition 1. *For all $\theta \in (0,1)$ and $\frac{2-\theta^2}{6-\theta^2} \leq \underline{k} < 1$, firm 1's profits are strictly ordered as*

$$\Pi_1^{DD} > \Pi_1^{DN} > \Pi_1^{NN}.$$

For firm 2, profits satisfy

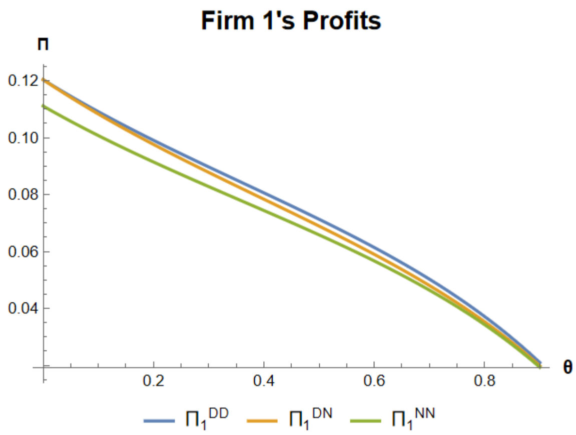
$$\Pi_2^{DD} > \Pi_2^{DN} = \Pi_2^{NN}.$$

Therefore, the Subgame Nash Equilibrium for the downstream market is for both firms to have access to consumer data.

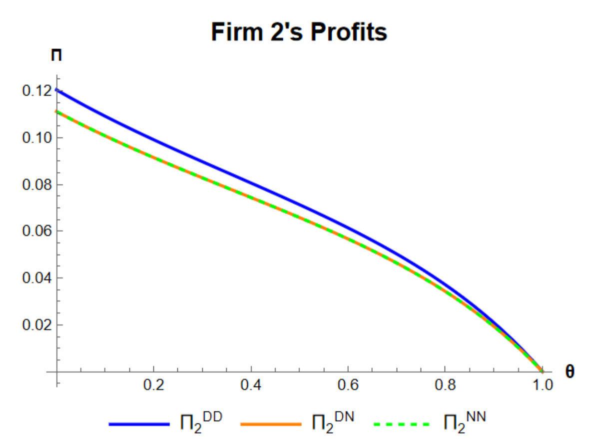
Proof. See Appendix A □

Both firms having access to consumer data is a dominant strategy, thereby resulting in this outcome in the downstream market game. Despite the seemingly contradictory nature of this phenomenon, this outcome is attributed to the pricing strategies employed by firm 2. In the DN case, firm 2 charges the average price, thereby expanding its market share among high-valuation consumers. Consequently, the unilateral purchase of consumer data becomes an inferior alternative for firm 1, compared to DD, since it is an attractive choice only for low-valuation consumers. when both firms obtain consumer data, they are able to charge high-valuation consumers at a higher rate and low-valuation consumers at a lower rate, thereby extracting a greater profit margin from the former. Figure 7 visually corroborates this result.

Note that when $\theta = 0$ or $\theta = 1$ firm 1's profits are the same. This is because at the first case products are fully separated therefore there is no immediate competition with firm 2. In the second case, products are practically the same, leading to the Bertrand's Paradox. However, at $\theta = 0$ for firm 2's profits in the DN case lie lower since they do not have access in consumers' data. In terms of total profits, consumer surplus and total welfare, the following figure illustrates a comparison of the results between the three regimes.

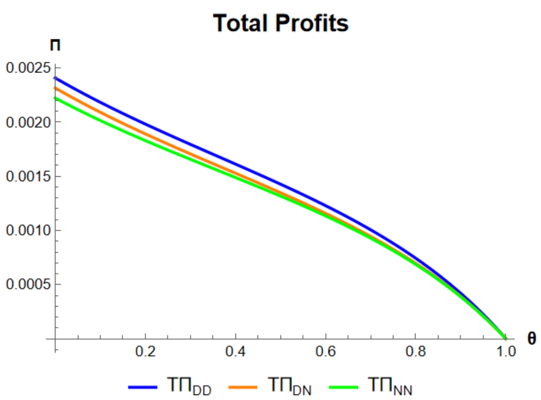


(a) Firm 1 Profit Comparison

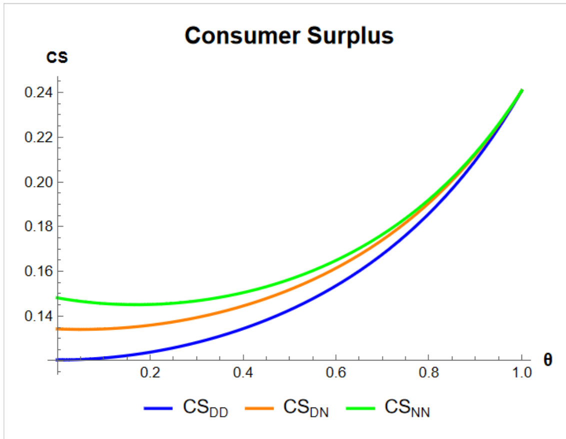


(b) Firm 2 Profit Comparison

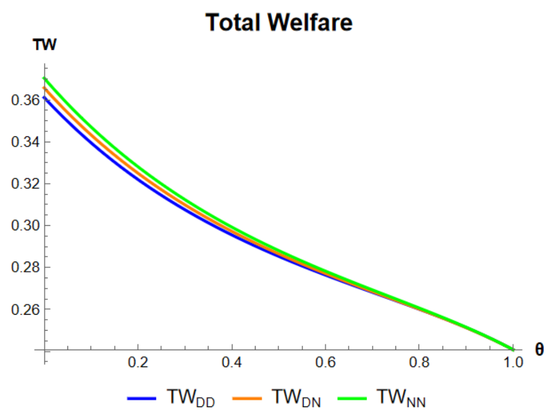
Figure 4: Firms 1 and 2 Profit Comparison



(a) Total Profit Comparison



(b) Consumer Surplus Comparison



(c) Total Welfare Comparison

Figure 5: Total Profits, Consumer Surplus and Total Welfare Comparison

Proposition 2. *For all $\theta \in (0,1)$ and $\frac{2-\theta^2}{6-\theta^2} \leq \underline{k} < 1$, total profits, consumer surplus and total welfare are ordered as:*

$$T\Pi_{DD} > T\Pi_{DN} > T\Pi_{NN}$$
$$TCS_{NN} > TCS_{DN} > TCS_{DD}$$
$$TW_{DD} > TW_{DN} > TW_{NN}$$

Proof. See Appendix A □

As a natural outcome, total profits under the DD regime are the highest, due to price discrimination and NN total profits are the lowest due to uniform pricing. Price discrimination generates the lowest consumer surplus, whereas the Bertrand competition yields the highest one. As previously mentioned, aggregate quantity does not increase with price discrimination and therefore, total welfare in both DN and DD lie below the NN case, with $TW_{DN} > TW_{DD}$

3.4.1 Upstream Firm

Let us now suppose that, the upstream firm charges a fixed cost to firm(s) for accessing/ purchasing the database. This cost is such that it extracts all of the firm(s) extra profit when they obtain the data, minus a small $\varepsilon > 0$, in order to make the companies buy the dataset. The profit of the upstream firm solely depends on the fixed price obtained by the downstream firms. If the data broker offers an exclusive contract to one of the downstream firms it will get $T_1 = \Pi_j^{DN} - \Pi_j^{NN} - \varepsilon$. In the case where the upstream firm can offer both firms access to the database, its profit is $2 \cdot (T_2 - \varepsilon)$, where $T_2 = \Pi_j^{DD} - \Pi_j^{NN}$, $j = 1, 2$. Therefore, the upstream firm's problem is the following[7]:

$$\Pi_U = max\{T_1, 2\,T_2\}$$

The following proposition shows the optimal decision for the data broker.

Proposition 3. *For all $\theta \in (0,1)$ and $\frac{2-\theta^2}{6-\theta^2} \leq \underline{k} < 1$, the upstream firm sells consumer data to both firms since in equilibrium $2\,T_2 > T_1$.*

Proof. If the upstream firm offers an exclusive contact it will only take $\Pi_1^{DN} - \Pi_1^{NN}$ which is less than $\Pi_j^{DD} - \Pi_j^{NN}$ (see Proposition 1). Therefore, the data broker will sell the dataset to both firms and gain $2 \cdot \Pi_j^{DD} - \Pi_j^{NN} \gg \Pi_1^{DN} - \Pi_1^{NN}\ \forall \theta \in (0,1)$ □

Figure 6 graphically illustrates the data broker's decision. One can see that only when products are perfect substitutes, the upstream firm is indifferent between selling to one or both firms. Of course, in this case, firms have the same profits in all cases (which are equal to 0) and therefore they are indifferent between buying or not buying the data. However, for all $0 < \theta < 1$, the optimal strategy for the upstream firm is to sell the data to both firms.

4 Conclusions

In this work, we employ a model of product differentiation and price discrimination in the presence of consumer data. Specifically, we examine a market structure comprising two downstream firms that engage in price competition, and an upstream firm that possesses consumer data. This upstream firm has the ability to sell this data to one or both of the downstream

[7] Here, we drop the ε, for simplicity

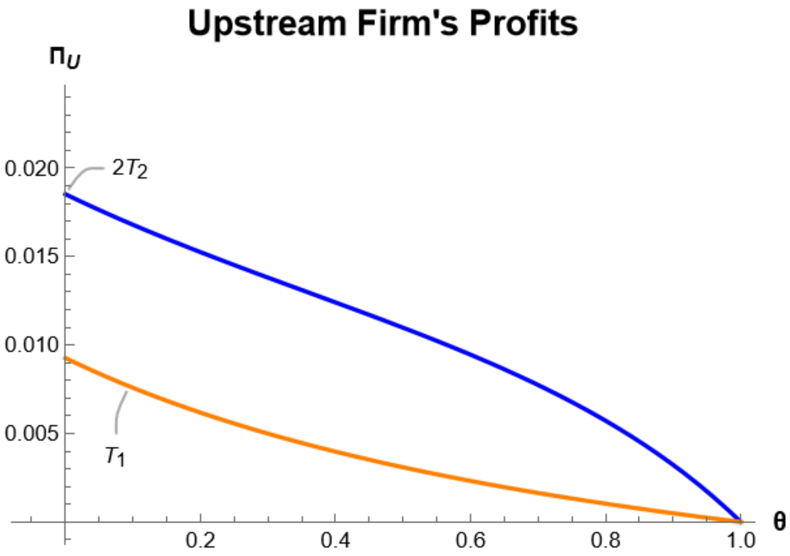


Figure 6: Upstream Firm Profits under Exclusive contract vs. Data Sharing

firms, thereby influencing the competitive landscape of the market.

Contrary to the findings of previous literature, our analysis reveals that the unique equilibrium in this market is characterized by both downstream firms acquiring the dataset from the upstream firm. Furthermore, we find that the upstream firm has a strong incentive to sell the data to both firms, as this allows it to maximize its profits. From a welfare perspective, our results indicate that total profits are maximized when both downstream firms acquire the dataset. However, this comes at the cost of minimized consumer surplus and welfare. In contrast, the Bertrand Competition scenario yields the highest levels of consumer surplus and welfare. This highlights the potential negative consequences of data-driven competition, where firms may prioritize profit maximization over consumer well-being.

Our findings have important implications for policymakers and regulators, who must balance the need to promote competition and innovation with the need to protect consumer interests. By understanding the complex interactions between firms and consumers in data-driven markets, we can develop more effective policies to promote welfare and prevent the exploitation of consumers.

Further extensions of this model that we are considering are (a) the introduction of a privacy cost, which allows consumers to endogenously choose for privacy, (b) the endogenous lower bound for positive residual demand, and (c) the distribution of θ.

References

Acquisti, Alessandro, Curtis Taylor, and Liad Wagman (2016). "The Economics of Privacy". In: *Journal of Economic Literature* 54.2, pages 442–492. DOI: 10.1257/jel.54.2.442. URL: https://pubs.aeaweb.org/doi/10.1257/jel.54.2.442.

Acquisti, Alessandro and Hal R Varian (2005). "Conditioning prices on purchase history". In: *Marketing Science* 24.3, pages 367–381.

Bounie, David, Antoine Dubus, and Patrick Waelbroeck (2021). "Selling strategic information in digital competitive markets". In: *The RAND Journal of Economics* 52.2, pages 283–313.

Hannak, Aniko et al. (2014). "Measuring price discrimination and steering on e-commerce web sites". In: *Proceedings of the 2014 conference on internet measurement conference*, pages 305–318.

Ichihashi, Shota (2020). "Online privacy and information disclosure by consumers". In: *American Economic Review* 110.2, pages 569–595.

Mikians, Jakub et al. (2013). "Crowd-assisted search for price discrimination in e-commerce: First results". In: *Proceedings of the ninth ACM conference on Emerging networking experiments and technologies*, pages 1–6.

Montes, Rodrigo, Wilfried Sand-Zantman, and Tommaso Valletti (2019). "The Value of Personal Information in Online Markets with Endogenous Privacy". In: *Management Science* 65.3, pages 1342–1362. DOI: 10.1287/mnsc.2017.2989. URL: https://pubsonline.informs.org/doi/10.1287/mnsc.2017.2989.

Pino, Flavio (2022). "The microeconomics of data – a survey". In: *J. Ind. Bus. Econ.* 49.3, pages 635–665. DOI: 10.1007/s40812-022-00220-6. URL: https://link.springer.com/10.1007/s40812-022-00220-6.

Taylor, Curtis and Liad Wagman (2014). "Consumer privacy in oligopolistic markets: Winners, losers, and welfare". In: *International Journal of Industrial Organization* 34, pages 80–84.

Tirole, Jean (1988). *The theory of industrial organization*. MIT press.

Woollaston, Victoria (2014). *Internet shoppers beware: Websites change product prices depending on a customer's purchase history - and even where they live*. URL: https://www.dailymail.co.uk/sciencetech/article-2818613/Internet-shoppers-beware-Websites-change-product-prices-depending-customers-purchase-history-live.html.

A Appendix

Proof for $\Pi_1^{DN} > \Pi_2^{DN}$

Proof. Here, we are going to show that $\Pi_1^{DN} > \Pi_2^{DN}$. For this to be true it must be that:

$$\frac{\alpha^2(1-\theta)\left[(\theta^2-4\theta)\left(1-\underline{k}\right)^2+16\left(\underline{k}^2+\underline{k}+1\right)\right]}{48b(2-\theta)^2(\theta+1)} > \frac{\alpha^2(1-\theta)\left(\underline{k}+1\right)^2}{4b(2-\theta)^2(\theta+1)} \Leftrightarrow$$

$$16(\underline{k}^2+\underline{k}+1)+(1-\underline{k})^2(\theta^2-4\theta) > 12(1+\underline{k})^2 \Leftrightarrow$$

After some mathematical manipulations, we have that:

$$4 > \theta(4-\theta)$$

which is true since $\theta \in [0,1]$. Therefore, $\Pi_1^{DN} > \Pi_2^{DN}$ □

Proposition 1. *For all $\theta \in (0,1)$ and $\dfrac{2-\theta^2}{6-\theta^2} \leq \underline{k} < 1$, firm 1's profits are strictly ordered as*

$$\Pi_1^{DD} > \Pi_1^{DN} > \Pi_1^{NN}.$$

For firm 2, profits satisfy

$$\Pi_2^{DD} > \Pi_2^{DN} = \Pi_2^{NN}.$$

Therefore, the Subgame Nash Equilibrium for the downstream market is for both firms to have access to consumer data.

Proof. We have that $\Pi_1^{DD} = \Pi_2^{DD} = \dfrac{\alpha^2(1-\theta)\left(\underline{k}^2+\underline{k}+1\right)}{3b(2-\theta)^2(\theta+1)}$,
$\Pi_1^{DN} = \dfrac{\alpha^2(1-\theta)\left[\theta^2\left(1-\underline{k}\right)^2-4\theta\left(1-\underline{k}\right)^2+16\left(\underline{k}^2+\underline{k}+1\right)\right]}{48b(2-\theta)^2(\theta+1)}$, and $\Pi_2^{DN} = \Pi_2^{NN} = \Pi_1^{NN} = \dfrac{\alpha^2(1-\theta)\left(\underline{k}+1\right)^2}{4b(2-\theta)^2(\theta+1)}$. Suppose that

$$\Pi_1^{DD} > \Pi_1^{DN} \Leftrightarrow$$

$$\frac{\alpha^2(1-\theta)\left(\underline{k}^2+\underline{k}+1\right)}{3b(2-\theta)^2(\theta+1)} > \frac{\alpha^2(1-\theta)\left[\theta^2\left(1-\underline{k}\right)^2-4\theta\left(1-\underline{k}\right)^2+16\left(\underline{k}^2+\underline{k}+1\right)\right]}{48b(2-\theta)^2(\theta+1)} \Leftrightarrow$$

$$(1-k)^2\theta(\theta-4) < 0$$

which is true for $\theta \in (0,1)$. Therefore, $\Pi_1^{DD} > \Pi_1^{DN}$. Moreover, for $\Pi_1^{DD} > \Pi_1^{NN}$ to be true, it must be that:

$$\frac{\alpha^2(1-\theta)\left(\underline{k}^2+\underline{k}+1\right)}{3b(2-\theta)^2(\theta+1)} > \frac{\alpha^2(1-\theta)\left(\underline{k}+1\right)^2}{4b(2-\theta)^2(\theta+1)} \Leftrightarrow$$

$$4\left(\underline{k}^2+\underline{k}+1\right) > 3\left(\underline{k}+1\right)^2 \Leftrightarrow$$

$$(\underline{k}-1)^2 > 0$$

which is true $\forall\theta \in (0,1)$, $\underline{k} > 0$. Finally, since $\Pi_1^{DD} = \Pi_2^{DD} > \Pi_1^{NN} = \Pi_2^{NN} = \Pi_2^{DN}$ then $\Pi_2^{DD} > \Pi_2^{DN} = \Pi_2^{NN}$ and $\Pi_1^{DN} > \Pi_1^{NN}$. □

Proposition 2. *For all $\theta \in (0,1)$ and $\dfrac{2-\theta^2}{6-\theta^2} \leq \underline{k} < 1$, total profits, consumer surplus and total welfare are ordered as:*

$$T\Pi_{DD} > T\Pi_{DN} > T\Pi_{NN}$$

$$TCS_{NN} > TCS_{DN} > TCS_{DD}$$

$$TW_{DD} > TW_{DN} > TW_{NN}$$

Proof. Since $\Pi_1^{DD} > \Pi_1^{DN} > \Pi_1^{NN}$ and $\Pi_2^{DD} > \Pi_2^{DN} = \Pi_2^{NN}$, we have that: $\Pi_1^{DD} + \Pi_2^{DD} > \Pi_1^{DN} + \Pi_2^{DN}$ and from $\Pi_1^{DN} > \Pi_1^{NN}$ we add to both sides Π_2^{DN} to have $\Pi_1^{DN} + \Pi_2^{DN} > \Pi_1^{NN} + \Pi_2^{DN}$. We also have that $\Pi_2^{DN} = \Pi_2^{NN}$ which ultimately leads to $\Pi_1^{DN} + \Pi_2^{DN} > \Pi_1^{NN} + \Pi_2^{NN}$.

For the consumer surplus we have that $TCS_{NN} > TCS_{DN} > TCS_{DD}$. Let us prove that $TCS_{NN} > TCS_{DN}$. For this to be true it must be that:

$$\frac{\alpha^2\left[\left(\theta^2-4\theta\right)\left(1-\underline{k}\right)^2+7\underline{k}^2-2\underline{k}+7\right]}{12b(2-\theta)^2(\theta+1)} > \frac{\alpha^2\left[\left(3\theta^3-7\theta^2-8\theta\right)\left(1-\underline{k}\right)^2+44\underline{k}^2+8\underline{k}+44\right]}{96b(2-\theta)^2(\theta+1)} \Leftrightarrow$$

$$8\left(\left(\theta^2-4\theta\right)\left(\underline{k}-1\right)^2+7\underline{k}^2-2\underline{k}+7\right) > \left(3\theta^3-7\theta^2-8\theta\right)\left(\underline{k}-1\right)^2+44\underline{k}^2+8\underline{k}+44 \Leftrightarrow$$

$$3(2-\theta)^2(1-\theta)\left(1-\underline{k}\right)^2 > 0$$

which is true. Therefore, $TCS_{NN} > TCS_{DN}$. Now we will check whether $TCS_{DN} > TCS_{DD}$. It must be that:

$$\frac{\alpha^2\left[\left(3\theta^3-7\theta^2-8\theta\right)\left(1-\underline{k}\right)^2+44\underline{k}^2+8\underline{k}+44\right]}{96b(2-\theta)^2(\theta+1)} > \frac{\alpha^2\left(\underline{k}^2+\underline{k}+1\right)}{3b(2-\theta)^2(\theta+1)} \Leftrightarrow$$

$$32\left(\underline{k}^2+\underline{k}+1\right) < 3\theta^3\left(\underline{k}-1\right)^2-7\theta^2\left(\underline{k}-1\right)^2-8\theta\left(\underline{k}-1\right)^2+44\underline{k}^2+8\underline{k}+44 \Leftrightarrow$$

$$(1-\theta)[\theta(4-3\theta)+12]\,(1-\underline{k})^2 > 0$$

which is true $\forall\theta \in (0,1)$ Therefore, $TCS_{NN} > TCS_{DN} > TCS_{DD}$. Finally, we have $TCS_{NN} > TCS_{DN} > TCS_{DD}$ and $T\Pi_{DD} > T\Pi_{DN} > T\Pi_{NN}$. Since $TW = TCS - T\Pi$, we can multiply the second inequality with -1 to get $-T\Pi_{DD} < -T\Pi_{DN} < -T\Pi_{NN}$. By adding the CS inequality and the latter we have that:

$$TCS_{DD} - T\Pi_{DD} < TCS_{DN} - T\Pi_{DN} < TCS_{NN} - T\Pi_{NN} \Leftrightarrow$$

$$TW_{NN} > TW_{DN} > TW_{DD}$$

$\square$