Data and the regulation of e-commerce: data sharing vs. dismantling

Claire Borsenberger[1], Helmuth Cremer[2], Denis Joram[3]
Jean-Marie Lozachmeur[4] and Estelle Malavolti[5]

April 2020

[1]DRAIE, Groupe La Poste
[2]TSE, University of Toulouse Capitole
[3]DRAIE, Groupe La Poste
[4]TSE, CNRS and University of Toulouse Capitole
[5]TSE, ENAC

Abstract

This paper considers an e-commerce market wherein a vertically integrated marketplace competes downstream with a single retailer and upstream with an independent parcel delivery operator. Because of the information collected by the marketplace on customers' habits and preferences, the integrated parcel delivery operator has lower delivery costs than its competitor. Products are differentiated according to the retailer and the parcel operator who delivers them. The representation of product differentiation is inspired by the Anderson, De Palma and Thisse (2002) discrete choice model.

We study several scenarios each representing a specific policy implemented to regulate the marketplace. The first one is a data sharing policy. The integrated marketplace has to share its information with the other delivery operator which in turn will lower this operator's cost of delivering the marketplace's product. The second one is vertical separation under which the parcel delivery operator previously owned and managed by the marketplace becomes independent. Finally we consider a full dismantlement scenario under which there is both vertical and horizontal separation.

We show that the optimal policy is either complete dismantlement or data sharing. The relative impacts on consumer surplus and total welfare of these two options involve a tradeoff between the increased competition implied by complete dismantling and the data related delivery cost advantage achieved under data sharing. When this cost advantage is small, completely dismantling dominates, while data sharing is the best policy when the cost advantage is large.
Keywords: E-commerce, delivery operators, vertical integration, platform regulation, data sharing, dismantling
JEL Codes: L42, L81, L87.

1 Introduction

The economic and societal role of digital platforms has been a hotly debated topic. They are under close scrutiny by European competition authorities for a while and their US counterparts have now followed suit. The subject is also receiving increasing attention in the media and in political circles. Each platform raises specific questions but the general themes are market power, the collection and (mis)use of personal data and related privacy issues, free speech and for some even their possible interference in the political process. Consequently the call for regulatory or competition policy intervention has become ever more pressing. Various reforms are considered including extreme solutions such a dismantlement of the platform.

The e-commerce sector, which has seen the increasing concentration of market power, is no exception. It has witnessed the emergence of marketplaces (a sort of horizontal integration of various independent retailers allowing the marketplace to offer a long tail of products) and a trend to vertical integration. In particular, the data collected by major platforms on their users (on the platforms' both side) provide them with a competitive edge over their competitors on all the markets they are involved in (retail but other parcel delivery in our case). On the demand side, it allows them to customize their search engines to customer profiles and use sophisticated pricing strategies. But it is also significant on the cost side as superior information allows an integrated marketplace to optimize its logistics and delivery network. This comes on top of traditional market power issues raised by horizontal and vertical integration.

In this paper we focus on the cost advantage associated with data collection. We study the equilibrium that emerges when a vertically integrated marketplace competes downstream with a single retailer and upstream with an independent parcel delivery operator. Because of the information collected by the marketplace, the integrated parcel delivery operator has lower delivery costs than its competitor. Products are differentiated according to the retailer and the parcel operator who delivers them. The representation of product differentiation is inspired by the Anderson, De Palma and Thisse (2002) discrete choice model and its application to the e-commerce sector follows

Borsenberger et al. (2020).

We study several scenarios each representing a specific policy implemented to regulate the marketplace. The first one is a data sharing policy. The integrated marketplace has to share its information with the other delivery operator which in turn will lower this operator's cost of delivering the marketplace's products. The second one is vertical separation under which the operator previously owned and managed by the marketplace becomes independent. Finally we consider a full dismantlement scenario under which there is both vertical and horizontal separation. The retailers which were previously affiliated with the marketplace now become independent.

The main conclusion we obtain is that the optimal policy is either complete dismantlement or data sharing. The relative impacts on consumer surplus and total welfare of these two options involve a tradeoff between the increased competition implied by complete dismantling and the data related delivery cost advantage achieved under data sharing. When this cost advantage is small, completely dismantling dominates, while data sharing is the best policy when the cost advantage is large. Vertical separation is never optimal. While it may or may not yield a larger welfare than the reference scenario it is always dominated by the two other policies.

2 The model

We consider an e-commerce sector with three retailers indexed $j = A$, M and B, and two delivery operators $i = 1, 2$. Initially, retailer A sells it "own" products but is also active as a marketplace which represents the exclusive outlet of retailer M. This assumption is the simplest way to represent the superior market power of the marketplace. Most significantly, it implies that when the marketplace is fully dismantled the total number of variants does not change.[1] Products are differentiated according to the retailer and the mode of delivery. Consequently there are initially six variants of the product (of which a total of four is sold by the marketplace).

Product differentiation is represented by the Anderson-De Palma-Thisse (1992) dis-

[1] Otherwise there would be a bias in the comparison across scenarios in favor of complete dismantlement.

crete choice model. A differentiated product is sold by downstream sellers A/M and B with marginal cost of k_j. It is shipped via differentiated upstream parcel delivery operators 1 and 2 (indexed by i) with marginal costs of c_{ij}. In the initial scenario operator 1 is owned and managed by A/M. Vertical integration gives the parcel delivery operator integrated with the marketplace superior information which is reflected by a lower marginal cost incurred to deliver products sold on the marketplace, than the cost incurred by the independent parcel delivery operator 2 to deliver the same products. Furthermore we assume that delivery operator 2 incurs the same marginal cost to deliver the products sold on the marketplace than the product sold by the independent retailer B. Consequently, we have $c_{1A/M} = \gamma^1 < c_{2A/M} = c_{2B} = \gamma^2$. For simplicity we assume $c_{1B} = c_{2B} = \gamma^2$: when it delivers for retailer B, delivery operator 1 does not benefit from superior information. Consequently, there are two relevant levels of marginal costs. A low level, γ^1, which applies when the integrated operator delivers the marketplace's parcels, and a higher level, γ^2, which applies to all other delivery flows.

There is a mass 1 of consumers. Consumer l derives utility

$$U_{ij}^l = b - p_{ij} + \varepsilon_{ij}^l$$

from consuming good ij where $j = A/M, B$ and $i = 1, 2$. The random variables ε_{ij}^l are identically and independently distributed across consumers and products with double exponential distribution over $\mathbb{R}$ with scale parameter σ.[2] We assume that the marketplace sells two variants (at the same price). Its demand is thus the sum of demands addressed to two retailers A and M. In the initial scenario there are thus six variants of the product, four of which are sold by the marketplace). However, consumers also have the option not to participate in the market. To model this we introduce an outside option as a seventh variant, indexed 00 with a given price p_{00}.

The parameter σ reflects the degree of product differentiation. When σ is small, the different variants are close substitutes and competition is intense.[3] When σ is large, each variant has roughly speaking a local monopoly and competition is not very intense.

[2] The distribution function of the double exponential distribution is $F(x) = \exp\left(-\exp\left(-\frac{x}{\sigma}\right)\right)$.

[3] When the distribution of x is given by $F(x) = \exp\left(-\exp\left(-\frac{x}{\sigma}\right)\right)$, a smaller σ means that there is a larger probability of x exceeding a given threshold. This can be interpreted as the products supplied being closer substitutes.

Consumers buy their preferred variant of the product if any. Consequently consumer l buys product ij when

$$U_{ij}^{l} \geq \max_{mn \neq ij} \{U_{mn}\}$$

It can be shown that the demand for good iA/M is then given by

$$D_{iA/M}(\mathbf{p}) = \frac{2\exp\left(-\frac{p_{iA/M}}{\sigma}\right)}{\sum_{i=1,2}\sum_{j=A/M,B}\exp\left(-\frac{p_{ij}}{\sigma}\right) + \exp\left(-\frac{p_{00}}{\sigma}\right)}, \tag{1}$$

while the demand for good iB is

$$D_{iB}(\mathbf{p}) = \frac{\exp\left(-\frac{p_{iB}}{\sigma}\right)}{\sum_{i=1,2}\sum_{j=A/M,B}\exp\left(-\frac{p_{ij}}{\sigma}\right) + \exp\left(-\frac{p_{00}}{\sigma}\right)}, \tag{2}$$

where $\mathbf{p} = \left(p_{1A/M}, p_{1B}, p_{2A/M}, p_{2B}\right)$ is the vector of consumer prices.

To understand (1) note that the marketplace sells two products: its own and that of the affiliated seller. We assume that these are sold at the same price. The price may vary according to the mode of delivery, though. One can thus think about the four variants as consisting of two pairs (one for each delivery operator) with the products in a given pair sold at the same price.[4]

The impact of prices on demand levels are given in Appendix A.1. The expressions show that a variant's market share is not surprisingly a decreasing function of its price. Furthermore, demand for any good increases if the price of one of the other variants increases. In other words, the variants are substitutes. The expressions also illustrate the role of the parameter σ. In particular the cross price effect is the larger the smaller is σ which is in line with our discussion above.

We consider four main scenarios. The reference case has already been sketched. The other scenarios study the implication of specific regulatory measures which are effectively discussed in practice. The first of these requires the integrated marketplace/delivery operator to share its data with the other actors. The second one consists in vertical separatism. Finally, the third one considers a more drastic reform where the integrated firm is dismantled both horizontally and vertically.

[4]Even if they would be allowed to differ, these price would be equal in equilibrium by symmetry. Consequently our assumption is not neccessary, but it is convenient for it simplifies notation.

3 The scenarios

We shall now present the different scenarios. For each of them we define the underlying game and specifically its timing. In all scenarios the game is sequential where delivery rates are set first followed by consumer prices. However, vertical separation or total dismantlement affect the number of players and the strategy space.

We start with the reference scenario, R, which is essentially the game already presented in the previous section. We then consider Scenario S, with data sharing. The structure of the game under data sharing is the same as under the reference scenario but the cost of the independent operator for delivering the marketplace's products is reduced (possibly to the same level as for the integrated operator). Scenario VS represents a more drastic departure: there are now two independent delivery operators which has a significant impact on the players' payoffs in Stage 1 and adds a new strategic variable, namely $t_{1A/M}$, the delivery rate set by delivery operator 1 to deliver the marketplace's product. Intuitively one can expect that vertical separation reduces market power but comes at the expense of introducing extra double marginalization. In other words, the traditional effects of vertical (dis)integration can be expected to be relevant.

3.1 Reference scenario: *R*

We consider a sequential game where delivery rates are set first followed by prices. We determine a subgame perfect equilibrium which means that in Stage 1, delivery operators anticipate the price equilibrium induced in Stage 2. We impose no *a priori* vertical restraints such as bundling and foreclosure, but these may appear endogenously in equilibrium when the relevant transaction (demand for the variant) is zero.

Let $\mathbf{t}^R = \left(t_{1B}, t_{2A/M}, t_{2B}\right)$ denote the vector of the delivery rates relevant in this scenario. The timing of the game is as follow. In Stage 1, the integrated firm chooses t_{1B} to maximize

$$\begin{aligned}\Pi^R_{1A/M} = &\left(p_{1A/M} - c_{1A/M}\right) D_{1A/M}\left(\widehat{\mathbf{p}}^R\right) + \left(p_{2A/M} - t_{2A/M}\right) D_{2A/M}\left(\widehat{\mathbf{p}}^R\right) \\ &+ \left(t_{1B} - c_{1B}\right) D_{1B}\left(\widehat{\mathbf{p}}^R\right), \end{aligned} \tag{3}$$

while delivery operator 2 chooses t_{2B} and $t_{2A/M}$ to maximize

$$\Pi_2^R = (t_{2B} - c_{2B})\, D_{2B}(\widehat{\mathbf{p}}^R) + \left(t_{2A/M} - c_{2A/M}\right) D_{2A/M}(\widehat{\mathbf{p}}^R). \tag{4}$$

Recall that $c_{1A/M} = \gamma^1 < c_{2A/M} = c_{2B} = c_{1B} = \gamma^2$ so that there are two relevant levels of marginal costs. A low level, γ^1, which applies when the integrated operator delivers the marketplace's parcels, and a higher level, γ^2, which applies to all other delivery flows. Observe that at this stage prices are determined by the induced second stage equilibrium. Consequently, $\widehat{\mathbf{p}}^R$ is a function of $\mathbf{t}^R = \left(t_{1B}, t_{2A/M}, t_{2B}\right)$. In Stage 2, the integrated firm sets $p_{1A/M}$ and $p_{2A/M}$ in order to maximize $\pi_{1A/M}^R$ given by

$$\pi_{1A/M}^R = \left(p_{1A/M} - c_{1A/M}\right) D_{1A/M}\left(\mathbf{p}^R\right) + \left(p_{2A/M} - t_{2A/M}\right) D_{2A/M}\left(\mathbf{p}^R\right) + \left(t_{1B} - c_{1B}\right) D_{1B}\left(\mathbf{p}^R\right), \tag{5}$$

which is the same expression as (3), except that delivery rates are now given. Consumer prices are now decision variables. Retailer B simultaneously sets its prices p_{1B} and p_{2B} to maximize

$$\pi_B^R = \sum_{i=1,2} \left(p_{iB} - t_{iB}\right) D_{iB}\left(\mathbf{p}^R\right).$$

3.2 Data sharing: *S*

The marketplace is now required to share its data with the delivery operator 2. Consequently the cost for this operator of delivering variant $2A/M$ is now given by $\widetilde{\gamma}^2$, with $c_{2B} > \widetilde{\gamma}^2 \geq c_{1A/M}$. The timing of the game is the same as in scenario R, but the profit function of operator 2 changes.

In Stage 1, the integrated firm chooses t_{1B} to maximize

$$\begin{aligned}\Pi_{1A/M}^S = {} & \left(p_{1A/M} - c_{1A/M}\right) D_{1A/M}\left(\widehat{\mathbf{p}}^S\right) + \left(p_{2A/M} - t_{2A/M}\right) D_{2A/M}\left(\widehat{\mathbf{p}}^S\right) \\ & + \left(t_{1B} - c_{1B}\right) D_{1B}\left(\widehat{\mathbf{p}}^S\right),\end{aligned} \tag{6}$$

while delivery operator 2 chooses t_{2B} and $t_{2A/M}$ to maximize

$$\Pi_2^S = (t_{2B} - c_{2B})\, D_{2B}(\widehat{\mathbf{p}}^S) + \left(t_{2A/M} - \widetilde{\gamma}^2\right) D_{2A/M}(\widehat{\mathbf{p}}^S). \tag{7}$$

At this stage prices are determined by the induced second stage equilibrium. Consequently, $\widehat{\mathbf{p}}^S$ is a function of $\mathbf{t}^S = \left(t_{1B}, t_{2A/M}, t_{2B}\right)$.

In Stage 2, the integrated firm sets $p_{1A/M}$ and $p_{2A/M}$ in order to maximize $\pi^S_{1A/M}$ given by

$$\pi^S_{1A/M} = \left(p_{1A/M} - c_{1A/M}\right) D_{1A/M}\left(\mathbf{p}^S\right) + \left(p_{2A/M} - t_{2A/M}\right) D_{2A/M}\left(\mathbf{p}^S\right) + \left(t_{1B} - c_{1B}\right) D_{1B}\left(\mathbf{p}^S\right), \tag{8}$$

which is the same expression as (6), except that delivery rates are now given. Retailer B simultaneously sets its prices p_{1B} and p_{2B} to maximize

$$\pi^S_B = \sum_{i=1,2} \left(p_{iB} - t_{iB}\right) D_{iB}\left(\mathbf{p}^S\right).$$

Note that since the second stage is the same in scenarios R and S, we have $\widehat{\mathbf{p}}^R\left(t_{1B}, t_{2A/M}, t_{2B}\right) = \widehat{\mathbf{p}}^S\left(t_{1B}, t_{2A/M}, t_{2B}\right)$. However, the first stage objectives for operator 2, (6) and (7) differ. Consequently the solutions will differ unless $D_{2A/M} = 0$ in both scenarios.

3.3 Vertical separation: *VS*

This scenario is similar, except for the asymmetry it involves, to the reference scenario considered by Borsenberger *et al.* (2020). It differs from scenario R in two ways. First, there is no longer vertical integration between A/M and operator 1. Second, the separation removes the cost advantage of operator 1 when delivering product variant $1A/M$.

The timing of the game is as follows. In a first stage delivery operators $i = 1, 2$ simultaneously set rates prices $t_{iA/M}$ and t_{iB} for retailers A/M and B respectively. Their profit are given by

$$\Pi^{VS}_i = \sum_j \left(t_{ij} - \gamma^2\right) D_{ij}\left(\widehat{\mathbf{p}}^{VS}\right), \; i = 1, 2. \tag{9}$$

Note that the vector of delivery rates now has four arguments $\mathbf{t}^{VS} = \left(t_{1A/M}, t_{1B}, t_{2A/M}, t_{2B}\right)$. In stage 2, retailers $j = A/M, B$ simultaneously set their prices p_{1j} and p_{2j} by taking as given the delivery rates. Their profit are given by

$$\pi^{VS}_j = \sum_i \left(p_{ij} - t_{ij}\right) D_{ij}\left(\mathbf{p}^{VS}\right), \; j = 1, 2. \tag{10}$$

3.4 Complete dismantling: *CD*

Now the activities of retailers A and M (previously grouped into the marketplace) are separated. As the result there are now three independent retailers A, M and B. This does not affect the total number of variants but both the consumer price vector and the delivery rate vector now have six dimensions: $\mathbf{t}^{VS} = (t_{1A}, t_{1M}, t_{1B}, t_{2A}, t_{2M}, t_{2B})$ and $\mathbf{p}^{VS} = (p_{1A}, p_{1M}, p_{1B}, p_{2A}, p_{2M}, p_{2B})$. This is because retailers A and M can charge different prices and may have to pay different delivery charges.

In a first stage delivery operators $i = 1, 2$ simultaneously set rates t_{iA}, t_{iM} and t_{iB} for retailers A, M and B respectively. Their profit are given by

$$\Pi_i^{CD} = \sum_j \left(t_{ij} - \gamma^2\right) D_{ij} \left(\widehat{\mathbf{p}}^{CD}\right), \; i = 1, 2. \tag{11}$$

Compared to expression (9) the sum now has an extra term. Furthermore the induced second stage prices $\widehat{\mathbf{p}}^{VS}$ have different expressions.

In stage 2, retailers $j = A, M, B$ simultaneously set their prices p_{1j} and p_{2j} by taking as given the delivery rates. Their profit are given by

$$\pi_j^{VS} = \sum_i \left(p_{ij} - t_{ij}\right) D_{ij} \left(\mathbf{p}^{VS}\right), \; j = 1, 2. \tag{12}$$

4 Numerical Results

When a scenario implies symmetric retailers and operators, the model can be solved analytically but even then the expressions are not very telling; see Anderson et al. (1992). Among the scenarios defined in the previous section the only symmetric one is *CD.* All others involve some asymmetry and in these cases, obtaining analytical closed form solutions would be at best very tedious. However, the model has only few parameters so that numerical solutions are very informative. Note that the constant b has no impact on the results and can be fixed arbitrarily.[5] We set $b = 15$ in all our scenarios. Furthermore the absolute levels of costs are not relevant; one of the cost

[5] Setting b sufficiently large ensures that utilities are positive. However when the outside option is introduced via an extra variant with a given price rather than a constant utility level, this is of no relevance.

levels can be normalized at one without loss of generality. Consequently we set $\gamma^2 = 1$. This leaves us with four relevant parameters, namely σ, γ^1, $\tilde{\gamma}^2$ and p_{00}. Recall that σ reflects the degree of product differentiation; when it is small, the different variants are close substitutes and competition is intense. With γ^2 normalized at one, the parameter γ^1 measures cost of the integrated delivery operator as a proportion of that of the independent delivery operator. The lower is γ^1 the larger is the cost advantage implied by the data available to the integrated delivery operator. Similarly $\tilde{\gamma}^2$ measures the independent delivery operator's cost of delivering a variant sold by the marketplace under data sharing and relative to its original cost (absent of data sharing). Finally p_{00} is the price of the outside option relative to the cost of an independent delivery operator (which we have normalized at 1).

In our setting, it turns out that the crucial parameter is γ^1 and thus the cost advantage that data provides to the integrated delivery operator. Depending on the level of γ^1 two patterns of results emerge in particular concerning the most appropriate regulatory policy. We show this by considering two baseline scenarios: one with a relatively large level of γ^1 (small cost advantage) and one with a smaller level of γ^1 (large cost advantage). These scenarios reveal our main results and illustrate the underlying intuition. They are followed by a number of variants with different levels of the crucial parameters which show that the results are robust.

4.1 Baseline scenarios

In both of these scenarios we set $\sigma = 1$ and $p_{00} = 6 = 6\gamma^2$. As will become clear from the results these values ensure that competition intensity is rather large (relatively low σ) and the outside option sufficiently expensive (making it less attractive) so that a large share of the market is covered. In both cases in the reference scenario more than 95% of consumers buy one of the 6 variants.

4.1.1 Small cost advantage: $\gamma^1 = 0.9$

Recall that the independent delivery operator's cost is normalized at $\gamma^2 = 1$. We consider two possible scenarios under data sharing. In scenario S_1 we have $\tilde{\gamma}^2 = 0.9$

so that the independent delivery operator's cost when delivering a product sold by the marketplace become equal to that of the integrated delivery operator. In scenario S_2 it remains larger with $\tilde{\gamma}^2 = 0.95$.

The equilibria obtained in the different scenarios are presented in Table 1.[6] The results show that data sharing under the two considered assumptions regarding its impact on cost (S_1 and S_2) has no significant impact on consumer surplus but increases total welfare.[7] Not surprisingly those effects are more significant when data sharing results in full cost matching than when the cost of the independent delivery operator remains larger than that of the integrated delivery operator. The independent delivery operator benefits while the profit of the independent retailer decreases. Vertical separation decreases consumer surplus because double marginalization leads to a price increase of the marketplace's variants. This also allows retailer B to increase its prices and to realize a larger profit. Total surplus, on the other hand, increases compared to the previous scenarios. The examples below, however, demonstrate that this is not a robust result as total surplus under vertical separation may even be lower than in the reference scenario.

Finally, complete dismantling dominates all other scenarios both from the perspective of consumers and that of total welfare. The fact that this policy performs better than VS does not come as a surprise because costs are not affected while competition becomes more intense. One can expect that this is a robust result and this is confirmed in all the examples presented below. The comparison with S on the other hand is less trivial. Increased competition now comes at the expense of and increase in delivery costs because delivery operators no longer benefit from the data advantage (directly or via data sharing). In this scenario the cost advantage is rather small which explains that the competition effect dominates. These conflicting effects are confirmed by the following scenario.

[6] A $*$ in a cell means that the variable is not relevant in that scenario.

[7] Consumer surplus increases slightly but this is not apparent in the table where only two digits are displayed for the sake of readability.

Scenario	R	S_1	S_2	VS	CD
$p_{1A/M}$	3.78	3.78	3.78	5.05	4.41
$p_{2A/M}$	4.40	4.35	4.37	5.05	4.41
p_{1B}	5.37	5.37	5.37	4.61	4.41
p_{2B}	4.20	4.21	4.21	4.61	4.41
p_{1M}	$*$	$*$	$*$	$*$	4.41
p_{2M}	$*$	$*$	$*$	$*$	4.41
t_{1A}	$*$	$*$	$*$	2.91	2.94
$t_{2A/M}$	3.24	3.19	3.22	2.91	2.94
t_{1B}	4.10	4.10	4.10	2.90	2.94
t_{2B}	2.93	2.95	2.94	2.90	2.94
t_{1M}	$*$	$*$	$*$	$*$	2.94
t_{2M}	$*$	$*$	$*$	$*$	2.94
$\pi_{1A/M}$	1.88	1.88	1.88	0.90	0.93
π_2	0.91	0.94	0.92	0.90	0.93
π_A	$*$	$*$	$*$	1.14	0.47
π_B	0.27	0.26	0.26	0.71	0.47
π_M	$*$	$*$	$*$	$*$	0.47
$d_{1A/M}$	0.49	0.48	0.49	0.26	0.16
$d_{2A/M}$	0.26	0.27	0.27	0.26	0.16
d_{1B}	0.05	0.05	0.05	0.20	0.16
d_{2B}	0.16	0.15	0.16	0.20	0.16
d_{1M}	$*$	$*$	$*$	$*$	0.16
d_{2M}	$*$	$*$	$*$	$*$	0.16
CS	12.17	12.17	12.17	11.64	12.41
TS	15.24	15.27	15.25	15.31	15.71

Table 1: Baseline scenario with small cost advantage.

Scenario	R	S_1	S_2	VS	CD
$p_{1A/M}$	3.33	3.29	3.31	5.05	4.41
$p_{2A/M}$	3.97	3.75	3.86	5.05	4.41
p_{1B}	5.40	5.38	5.39	4.61	4.41
p_{2B}	4.04	4.15	4.09	4.61	4.41
p_{1M}	*	*	*	*	4.41
p_{2M}	*	*	*	*	4.41
t_{1A}	*	*	*	2.91	2.93
$t_{2A/M}$	3.24	2.98	3.10	2.91	2.93
t_{1B}	4.20	4.22	4.21	2.90	2.93
t_{2B}	2.85	2.98	2.91	2.90	2.93
t_{1M}	*	*	*	*	2.93
t_{2M}	*	*	*	*	2.93
$\pi_{1A/M}$	1.83	1.79	1.81	0.90	0.93
π_2	0.87	1.02	0.94	0.90	0.93
π_A	*	*	*	1.14	0.47
π_B	0.19	0.16	0.18	0.70	0.47
π_M	*	*	*	*	0.47
$d_{1A/M}$	0.53	0.51	0.52	0.26	0.16
$d_{2A/M}$	0.28	0.32	0.30	0.26	0.16
d_{1B}	0.03	0.03	0.03	0.20	0.16
d_{2B}	0.13	0.10	0.12	0.20	0.16
d_{1M}	*	*	*	*	0.16
d_{2M}	*	*	*	*	0.16
CS	12.92	12.96	12.94	11.64	12.41
TS	15.83	15.95	15.88	15.31	15.71

Table 2: Baseline scenario with large cost advantage.

4.1.2 Large cost advantage: $\gamma^1 = 0.5$

We now consider the case where the data related delivery cost advantage is more significant by assuming that $\gamma^1 = 0.5$. Once again two possible scenarios for data sharing are considered. In S_1 we have $\tilde{\gamma}^2 = 0.5$ so that the cost advantage is fully matched by the independent delivery operator. In S_2 we have $\tilde{\gamma}^2 = 0.75$ so that its cost decreases but remains larger than that of the integrated delivery operator. The results are shown in Table 2.

One notices that data sharing now has a more significant (though still small) impact on consumer surplus. However, the main interest of this example lies in the comparison between S and CD. The cost delivery cost advantage implied by the data (whether direct or shared) is now so significant that it outweighs the increased competition intensity brought about by total dismantling. In this case data sharing is the best policy.

4.2 Robustness checks

We now present a number of examples with other values of the relevant parameters. They show that while some of the observed effects on prices, delivery rates or profits are specific to the considered examples, the main conclusions appear to be robust. To be precise vertical separation is never the optimal policy. A regulating authority concerned with either consumer or total welfare should implement either data sharing or full dismantling. Which of these policies is determined by a tradeoff between delivery costs and competition intensity.

4.2.1 Larger scale factor implying lower competition intensity

Table 3 and 4 illustrate the two relevant cases when $\sigma = 1.5$. The other parameters are the same as before and so are the two scenarios regarding the data related delivery cost advantage.

Specifically, in Table 3 we have $\gamma^1 = 0.9$ and with two data sharing scenarios obtained for $\tilde{\gamma}^2 = 0.9$ in scenario S_1 and $\tilde{\gamma}^2 = 0.95$ in scenario S_2. This is the case where the cost advantage is small so that the competition effect dominates and implies the complete dismantling is the best policy. This shows that the main conclusion obtained from Table 1 remains valid even when competition intensity is smaller - a fact which can be expected to mitigate the positive competition effect associated with complete dismantling. In Table 4 we have $\gamma^1 = 0.5$ along with $\tilde{\gamma}^2 = 0.5$ in scenario S_1 and $\tilde{\gamma}^2 = 0.75$ in scenario S_2. Now the cost effect is again dominating and data sharing is best policy. In all scenarios presented in these two tables, vertical separation is the worst policy option and it even reduces social and consumer surplus compared to the reference scenario. Compared to the scenarios presented in Table 1, vertical separation thus performes worse here. This is in line with intuition: as product differentiation

Scenario	R	S_1	S_2	VS	CD
$p_{1A/M}$	4.87	4.88	4.87	6.52	5.79
$p_{2A/M}$	5.86	5.82	5.84	6.52	5.79
p_{1B}	6.92	6.92	6.92	5.93	5.79
p_{2B}	5.54	5.55	5.55	5.93	5.79
p_{1M}	$*$	$*$	$*$	$*$	5.79
p_{2M}	$*$	$*$	$*$	$*$	5.79
t_{1A}	$*$	$*$	$*$	3.65	3.68
$t_{2A/M}$	4.00	3.94	3.97	3.65	3.68
t_{1B}	5.04	5.04	5.04	3.61	3.68
t_{2B}	3.66	3.68	3.67	3.61	3.68
t_{1M}	$*$	$*$	$*$	$*$	3.68
t_{2M}	$*$	$*$	$*$	$*$	3.68
π_1	2.47	2.48	2.47	1.09	1.17
π_2	1.09	1.11	1.10	1.09	1.17
π_A	$*$	$*$	$*$	1.36	0.61
π_B	0.38	0.37	0.38	0.82	0.61
π_M	$*$	$*$	$*$	$*$	0.61
$d_{1A/M}$	0.45	0.45	0.45	0.23	0.14
$d_{2A/M}$	0.23	0.24	0.23	0.23	0.14
d_{1B}	0.05	0.05	0.05	0.17	0.14
d_{2B}	0.14	0.14	0.14	0.17	0.14
d_{1M}	$*$	$*$	$*$	$*$	0.14
d_{2M}	$*$	$*$	$*$	$*$	0.14
CS	11.82	11.83	11.83	11.25	12.09
TS	15.78	15.80	15.79	15.63	16.28

Table 3: Lower competition intensity and small cost advantage: $\gamma^1 = 0.9$ with $\tilde{\gamma}^2 = 0.9$ in scenario S_1 and $\tilde{\gamma}^2 = 0.95$ in scenario S_2.

becomes more significant, the benefits of increased competition are small and do not outweigh negative impact of double marginalization together with the loss of the data related cost advantage.

4.2.2 Large scale factor and more attractive outside option

We now consider an even larger level of $\sigma = 2.5$ together with a smaller level of the price of the outside option $p_{00} = 4 = 4\gamma^2$. Table 5 presents the results for the case where the cost advantage is small, while Table 6 is obtained for the larger cost advantage.

Scenario	R	S_1	S_2	VS	CD
$p_{1A/M}$	5.04	5.10	5.07	6.52	5.79
$p_{2A/M}$	7.20	6.86	7.02	6.52	5.79
p_{1B}	7.23	7.28	7.25	5.93	5.79
p_{2B}	5.40	5.43	5.42	5.93	5.79
p_{1M}	$*$	$*$	$*$	$*$	5.79
p_{2M}	$*$	$*$	$*$	$*$	5.79
t_{1A}	$*$	$*$	$*$	3.65	3.68
t_{2A}	2.66	2.25	2.45	3.65	3.68
t_{1B}	5.22	5.29	5.25	3.61	3.68
t_{2B}	3.40	3.44	3.42	3.61	3.68
t_{1M}	$*$	$*$	$*$	$*$	3.68
t_{2M}	$*$	$*$	$*$	$*$	3.68
$\pi_{1A/M}$	3.04	3.10	3.07	1.09	1.17
π_2	0.66	0.72	0.69	1.09	1.17
π_A	$*$	$*$	$*$	1.36	0.61
π_B	0.50	0.49	0.49	0.82	0.61
π_M	$*$	$*$	$*$	$*$	0.61
$d_{1A/M}$	0.49	0.47	0.48	0.23	0.14
$d_{2A/M}$	0.11	0.14	0.13	0.23	0.14
d_{1B}	0.05	0.05	0.05	0.17	0.14
d_{2B}	0.19	0.19	0.19	0.17	0.14
d_{1M}	$*$	$*$	$*$	$*$	0.14
d_{2M}	$*$	$*$	$*$	$*$	0.14
CS	12.68	12.68	12.68	11.25	12.09
TS	16.89	17.00	16.94	15.63	16.28

Table 4: Lower competition intensity and large cost advantage: $\gamma^1 = 0.5$ with $\tilde{\gamma}^2 = 0.5$ in scenario S_1 and $\tilde{\gamma}^2 = 0.75$ in scenario S_2.

Scenario	R	S	VS	CD
$p_{1A/M}$	5.97	5.98	8.38	7.72
$p_{2A/M}$	9.06	8.98	8.38	7.72
p_{1B}	8.93	8.94	7.73	7.72
p_{2B}	7.33	7.33	7.73	7.72
p_{1M}	$*$	$*$	$*$	7.72
p_{2M}	$*$	$*$	$*$	7.72
t_1	$*$	$*$	4.69	4.62
$t_{2A/M}$	3.98	3.90	4.69	4.62
t_{1B}	5.97	5.98	4.57	4.62
t_{2B}	4.37	4.37	4.57	4.62
t_{1M}	$*$	$*$	$*$	4.62
t_{2M}	$*$	$*$	$*$	4.62
$\pi_{1A/M}$	2.57	2.58	0.97	1.04
π_2	0.65	0.66	0.97	1.04
π_A	$*$	$*$	1.19	0.59
π_B	0.46	0.46	0.66	0.59
π_M	$*$	$*$	$*$	0.59
$d_{1A/M}$	0.35	0.35	0.16	0.09
$d_{2A/M}$	0.10	0.10	0.16	0.09
d_{1B}	0.05	0.05	0.10	0.09
d_{2B}	0.10	0.10	0.10	0.09
d_{1M}	$*$	$*$	$*$	0.09
d_{2M}	$*$	$*$	$*$	0.09
CS	12.92	12.92	12.46	13.14
TS	16.62	16.63	16.26	17.00

Table 5: Low competition intensity, small cost advantage and lower price of the outside option: $\sigma = 2.5$, $p_{00} = 4$, $\gamma^1 = 0.9$ with $\tilde{\gamma}^2 = 0.9$ in scenario S_1 and $\tilde{\gamma}^2 = 0.95$ in scenario S_2.

Scenario	R	S	VS	CD
$p_{1A/M}$	6.47	6.52	9.65	8.90
$p_{2A/M}$	10.50	10.12	9.65	8.90
p_{1B}	10.36	10.39	8.92	8.90
p_{2B}	8.43	8.46	8.92	8.90
p_{1M}	$*$	$*$	$*$	8.90
p_{2M}	$*$	$*$	$*$	8.90
t_{1A}	$*$	$*$	5.33	5.25
$t_{2A/M}$	4.52	4.10	5.33	5.25
t_{1B}	6.86	6.91	5.19	5.25
t_{2B}	4.93	4.97	5.19	5.25
t_{1M}	$*$	$*$	$*$	5.25
t_{2M}	$*$	$*$	$*$	5.25
$\pi_{1A/M}$	2.97	3.02	1.06	1.14
π_2	0.69	0.74	1.06	1.14
π_A	$*$	$*$	1.31	0.65
π_B	0.49	0.48	0.72	0.65
π_M	$*$	$*$	$*$	0.65
$d_{1A/M}$	0.35	0.34	0.15	0.08
$d_{2A/M}$	0.09	0.10	0.15	0.08
d_{1B}	0.04	0.04	0.09	0.08
d_{2B}	0.09	0.09	0.09	0.08
d_{1M}	$*$	$*$	$*$	0.08
d_{2M}	$*$	$*$	$*$	0.08
CS	14.13	14.14	12.57	13.32
TS	18.30	18.39	16.75	17.58

Table 6: Low competition intensity, large cost advantage and lower price of the outside option: $\sigma = 2.5$, $p_{00} = 4$, $\gamma^1 = 0.5$ with $\tilde{\gamma}^2 = 0.5$ in scenario S_1 and $\tilde{\gamma}^2 = 0.75$ in scenario S_2.

Not surprisingly, the large degree of product differentiation and the increased attractiveness of the outside option concur to bring about a significant drop in market coverage. Interestingly, this does not affect our main conclusions: complete dismantling is the best option with a small cost advantage while data sharing dominates when the cost advantage is more significant.

5 Conclusion

In this paper, we study several ways to regulate a vertically integrated marketplace in the parcel delivery sector, that benefits from a cost advantage in delivery due to data collected on consumers' habits and preferences through the retail activity. In particular, we compare three regulatory schemes: *(i)* imposing the integrated marketplace to share its information with the other delivery operator which in turn will lower this operator's cost of delivering the marketplace's product; *(ii)* imposing a vertical separation under which the delivery operator previously owned and managed by the marketplace becomes independent and no longer benefits from a cost advantage over its competitor; *(iii)* imposing a full dismantlement of the marketplace under which there is both vertical and horizontal separation (all retailers and delivery opertors, now independent, compete on their market segment). The main robust conclusion we obtain is that the optimal policy is either complete dismantlement or data sharing. The relative impacts on consumer surplus and total welfare of these two options involve a tradeoff between the increased competition implied by complete dismantling and the data related delivery cost advantage achieved under data sharing. When this cost advantage is small, completely dismantling dominates, while data sharing is the best policy when the cost advantage is large. Vertical separation is never optimal.

Our results are obtained in a simple and stylized model and have to be qualified accordingly. In particular we concentrate on delivery costs while in reality superior data also enhances the possibilities to practice sophisticated pricing schemes. One can expect that this makes data sharing an even more powerful regulatory tools. We have also neglected the possible "quality" advantage associated with marketplaces. As intermediary platform the marketplace provides tools to the different parties (producers/retailers and consumers/buyers) which simplify trading: online payment system, inventory management, authenticated information about the seller and/or the buyer, various warranties and more and more often integrated delivery services. Taking this effect into account when comparing data sharing and total dismantlement is likely to increase the number of cases in which data sharing is the best policy.

References

[1] Anderson, S., A De Palma and J.F. Thisse, "Discrete Choice Theory of Product Differentiation", MIT Press, 1992.

[2] Ben-Akiva, M. and S. Lerman, "Disaggregate Travel and Mobility Choice Models and Measures of Accessibility." In D.A. Hensher and P.R. Stopher (eds.), Behavioral Travel Modelling, London: Croon Helm, 654–679, 1979.

[3] Borsenberger, C., H. Cremer, D. Joram and J.-M. Lozachmeur, "Vertical integration in the e-commerce sector," in: Parcu, P., T. Brennan and V. Glass, eds, New business and Regulatory Strategies in the Postal Sector, 2018.

[4] Borsenberger, C., H. Cremer, D. Joram, J.-M. Lozachmeur and E. Malavolti, "Platform competition: market competition and pricing," in: Parcu, P., T. Brennan and V. Glass, eds, The changing postal environment: Market and Policy Innovation, 2020.

[5] EU Guidelines on Vertical Restraints, 2010.

[6] Petit, N. and D. Henry, 2010, "Vertical Restraints under EU Competition Law: Conceptual Foundations and Practical Framework", http://dx.doi.org/10.2139/ssrn.1724891

[7] Reisinger, M. and Tarantino, 2015, "Vertical integration, foreclosure, and productive efficiency", The Rand Journal of Economics, Vol. 46, No. 3 (Fall 2015), pp. 461-479.

[8] Rey, P. and J. Tirole, 2007, "A primer in foreclosure", Chapter 33 in Handbook of Industrial Organization, vol. 3, pp 2145-2220.

Appendix

A.1 Properties of demand functions

We have

$$\frac{\partial D_{ij}\left(\mathbf{p}\right)}{\partial p_{ij}} = \frac{-\frac{1}{\sigma}\exp\left(-\frac{p_{ij}}{\sigma}\right)\left(\sum_{i=1,2}\sum_{j=A,B}\exp\left(-\frac{p_{ij}}{\sigma}\right)+\exp\left(-\frac{p_{00}}{\sigma}\right)\right)+\frac{1}{\sigma}\exp\left(\frac{p_{ij}}{\sigma}\right)\exp\left(-\frac{p_{ij}}{\sigma}\right)}{\left(\sum_{i=1,2}\sum_{j=A,B}\exp\left(-\frac{p_{ij}}{\sigma}\right)+\exp\left(-\frac{p_{00}}{\sigma}\right)\right)^2}$$
$$= \frac{1}{\sigma}\left(-D_{ij}+D_{ij}^2\right)$$
$$= \frac{1}{\sigma}D_{ij}\left(D_{ij}-1\right) < 0,$$

and

$$\frac{\partial D_{ij}\left(\mathbf{p}\right)}{\partial p_{mn}} = \frac{\frac{1}{\sigma}\exp\left(-\frac{p_{ij}}{\sigma}\right)\exp\left(-\frac{p_{mn}}{\sigma}\right)}{\left(\sum_{i=1,2}\sum_{j=A,B}\exp\left(-\frac{p_{ij}}{\sigma}\right)+\exp\left(-\frac{p_{00}}{\sigma}\right)\right)^2}$$
$$= \frac{1}{\sigma}D_{ij}D_{mn} > 0.$$

Expected consumer surplus is given by

$$CS = \sigma\ln\left(\sum_{i=1,2}\sum_{j=A,B}\exp\left(\frac{b-p_{ij}}{\sigma}\right)+\exp\left(\frac{b-p_{00}}{\sigma}\right)\right),$$

see Ben-Akiva and Lerman (1979), p.114.