Segmentation and nonlinear pricing
in the postal sector[*]

Claire Borsenberger[†], Helmuth Cremer[‡], Philippe De Donder[§],
Sébastien Lécou[**] and Denis Joram[††]

June 2012

[†] Economist, European and National Regulation Division, Le Groupe La Poste.
[‡] Corresponding author: Toulouse of Economics (IDEI, GREMAQ and Iuf), helmuth.cremer@tse-fr.eu.
[§] Toulouse School of Economics (IDEI and GREMAQ-CNRS).
[**] Economist, European and National Regulation Division, Le Groupe La Poste.
[††] Chief Economist, Economist, European and National Regulation Division, Le Groupe La Poste.

1. Introduction

Network operators typically use quite sophisticated pricing policies. For instance, nonlinear pricing (second degree price discrimination) has a long tradition in the telecommunication and energy sectors. Similarly, in these and other sectors, prices are often differentiated according to market segments (third degree discrimination). Even though it is not the most prominent textbook example to illustrate such pricing policies, the postal sector is in fact no exception. Ramsey pricing is effectively a form of third degree price discrimination, based on price elasticities. In addition, most postal operators have volume discount programs (a form of nonlinear pricing). While regulators and competition authorities are often reluctant to accept these practices, economists tend to view differentiated pricing policies in a more positive way.[1] For a regulated operator, they are an effective way to cover fixed cost while mitigating distortions that would be associated with linear pricing. When the regulatory policy is well designed and the operator is welfare maximizing, sophisticated pricing policies can only be welfare enhancing. However, when the operator is profit maximizing, this is no longer necessarily true: the scenario where a profit-maximizing operator may engage in pricing policies that are detrimental to welfare can then no longer be ruled out. Still, the literature has shown that departures from linear pricing are often welfare enhancing, even when the operator is profit maximizing.[2]

The implications of differentiated pricing policies for the postal sector have received little attention in the literature. A notable exception is Crew and Kleindorfer (2011, 2012) who study the implications of volume discounts. They show that nonlinear pricing can contribute to the financial viability of an operator who faces market opening and intermodal competition (in particular from electronic media).

In this paper we study pricing policies, which combine market segmentation (third degree price discrimination) with nonlinear pricing (second degree price discrimination) - referred to as "segmented nonlinear pricing" (SNLP). More precisely, we assume that the operator can group customers into a certain number of categories on the basis of an exogenously observable characteristic at no cost. Under standard third degree discrimination, the operator charges linear prices that differ across categories. Under SNLP, the operator can use category-specific nonlinear tariffs. This is common practice for telecommunications (particularly mobile phone), in the advertising market and in the airline sector among other. It is also often referred to as tagging in particular in the public economics literature.[3] In the postal sector, SNLP has a number of potential applications and there already exist some examples where it is used in practice.

One of the pluses of SNLP is that it improves the design of volume discounts by targeting them to the categories of customers who are most likely to switch to electronic substitutes. In a context of decreasing mail volume, such pricing policies could mitigate this decrease and boost volumes. Consequently, this could become a crucial factor in safeguarding universal postal service.[4]

While our analysis provides a methodological contribution that goes beyond the postal

---

[1] Geradin (2011) and Eccles (2012) give a clear description of the key principles of EU and national competition laws on price discrimination in relation to the postal Directive.

[2] Willig (1978) provided an early discussion. Aguirre et al. (2010) is a rather comprehensive study of the welfare impact of third degree price discrimination

[3].This paper and a number of others, which are referenced in Cremer *et al.* (2010)studies the design of nonlinear tax schedules that can be conditioned on some observable characteristic (like age or gender). While this appears at first to be a very different question from the one we are considering here, the formal problem is very similar.

[4] A similar point is made by Crew and Kleindorfer (2012).

sector, we focus our attention on specific postal issues. This is reflected in our illustrative examples but also in our modeling assumptions. In particular, we include the feature that customers always have the option to use the public stamp rate (reservation utility is determined by a linear tariff and is thus type-specific).

The design of nonlinear tariffs is first and foremost an issue of asymmetric information. The operator does not observe the individual customer's willingness to pay. Under full information, the operator would induce an efficient allocation. This is because it can extract the entire surplus and thus chooses a policy to maximize it. Under asymmetric information, there is a trade-off between rent extraction and efficiency and quantities are distorted (typically downward) to mitigate informational rents. Under SNLP, the operator observes the customer's category, which provides some information about the distribution of his type (willingness to pay). More precisely, it can use the intra category distribution of types (rather than the overall distribution) to offer segment-specific tariffs.

We characterize the solution under SNLP in a simple four types, two categories setting and compare it to the standard NLP (nonlinear pricing) solution. We provide a set of sufficient conditions under which SNLP (used by a profit maximizing operator) is welfare improving. When these conditions are violated, a welfare improvement continues to be possible, but whether it materializes depends on the preference and cost structure. For either case, we use numerical examples to study the impact of the pricing schemes on the various customer groups and on the operator's profit.

We start by presenting some examples of SNLP schemes in various sectors (Section 2). Then we develop the basic model, wherein the willingness to pay in the two groups does not overlap and where the reservation utility is exogenous and the same for all individuals. This is the simplest setting and yields the most clear-cut results. We then study two extensions. In the first, we consider the case of overlapping groups (Subsection 1). In the second, we assume that a simple linear tariff (the stamp price) is available to all customers and determines their reservation utility levels (Subsection 2). We study how this specific feature of the postal sector affects our results. Finally we provide some illustrative numerical examples in Section 8.

2. Some examples of segmented nonlinear pricing schemes

Price discrimination is pervasive in all sectors of the economy. Firms use price discrimination to stimulate demand and increase their profit, but this practice could also be socially beneficial. In particular, in industries with huge fixed costs (and low marginal costs), such as the postal sector, price discrimination allows firms to spread fixed costs over a larger number of units and brings the market equilibrium closer to the social optimum. In sectors where volumes are shrinking, such as in the postal sector which is facing electronic substitution, price discrimination could make sense to maintain the minimum output level permitting to remain financially viable and safeguard the USO.

In practice, what economists call "second degree price discrimination" is particularly widespread: firms set a price per unit which varies with the number of units bought by the customer. This can be achieved through volume discounts or the adoption of a two-part tariff. Through these nonlinear pricing schemes, the supplier provides incentives for the customers to "self-select" into categories according to their preferences. Consequently the supplier can charge different prices to the different groups. Volume discounts are particularly well developed in the postal sector as well as in many other sectors producing intermediate goods. In practice, different rates may apply to differences in product quality as well as quantity. For example, airlines often offer multiple classes of seats on flights, such as first class and economy class. Similarly, most postal operators distinguish direct mail from transactional mail, services

with different transit time and quality of service. This form of discrimination based on quality is a mean to segment mail senders and to charge the transactional mail senders more than direct mailers, because the former have typically a less elastic demand.

When the suppliers are able to differentiate between consumer classes, pricing policy is referred to as "third degree discrimination": suppliers charge different prices to different groups of customers depending on their elasticity of demand. Price varies by attributes such as location, time or observed characteristics of the customer. Many movie theaters, leisure parks and tourist attractions have different admission prices according to the market segment; people are segmented into youth, student, adult, and senior groups. Each of these groups typically has a different demand function, based on different levels of disposable income.

These two types of discrimination are not mutually exclusive: a supplier may vary pricing by category of customers and offer volume discounts as well. Such practices are common to many economic sectors. A first example of the combination of second and third degree price discrimination can be found in the advertising market. For example, in France, major TV networks offer both rebates based on volume as well as "sectorial rebates" to advertisers in order to induce them to place more ads on TV programs. For example, *France Televisions* (French public channels) offer rebates to retailers and distance sellers (4 percent), to telecom providers (7 percent), to agricultural, gardening or DIY advertisers (7 percent). New advertisers are offered a rebate of 5 percent. Another discrimination criterion is the size of the advertisers: small ones (in terms of revenues) benefit from a discount of 10 percent.

Other examples of policies that combine types of price discrimination can be found in the travel industry. Airlines companies typically offer volume discounts to wholesalers, consolidators and tour operators, lower prices for customers who stay over on Saturday nights or buy their tickets in advance, and also roundtrip discounts. Airline price discrimination represents an attempt to charge business travelers more than leisure travelers (because the formers have a less elastic demand than the latter), but also to charge more for leisure travel during school holidays and for immediate departure than respectively travelling in low seasons and early booking through seasonal pricing and advance discounts.

In the telecommunication sector, prices generally decrease with the quantity of communication time consumed and vary according to the time of the day (different rates are applied for daytime, evening and week-end calls). In the energy sector, nonlinear tariffs are segmented by time, through peak and off-peak tariffs.

In the postal sector, this double discrimination appears to be less common (at least for the time being). Nevertheless, some examples can be found. In particular, several postal operators offer tariffs that combine second and third degree price discrimination for nonprofit organizations. In the US this specific category of mail senders has been benefitting since 1951 from reduced rates compared to other categories of senders. To qualify for these special terms, organizations must prove that they are both organized and operated for a qualifying primary purpose, that they are not organized for profit and that none of the organization's net income is intended to benefit any individual or private shareholder. Eight categories of nonprofit organizations are eligible for this rate: religious, educational, scientific, philanthropic, agricultural, labor, veterans and fraternal. In addition, these nonprofit organizations can benefit from volume discounts. This specific segmented nonlinear pricing scheme between profit and nonprofit organizations has since then been adopted by several postal operators (*bpost*, *Posten AB*, Australian Post, *Poste Italiane* among others). Recently, the French operator *La Poste* has launched an offer: since October 2011, nonprofit organizations, recognized as such by a legal act, are eligible to the offer "*Integral Utilite Publique*", characterized by a reduced rate. Recently several operators have also experimented with direct mail products that offer discounts when the mailing campaign is targeting new customers. These pricing scheme induce direct mailers to target new customers and to increase their mail demand, the lower return on

investment (sending mail to prospective advertising targets is likely to be less profitable than sending mail to already known targets) being partially compensated by lower expenditure on postage. Duncan *et al.* (2011) present examples which include "USPS Summer Sale" and "Royal Mail Sale", where postal operators propose short-term discounts and special offers. These authors also argue that sectorial price discrimination ought to be considered because underlying purposes, costs and revenues of the campaigns vary between sectors.

We now turn to a basic model to study the impact of various pricing schemes, notably the combination of second and third price discrimination practices (through segmented nonlinear pricing schemes), on welfare.

3. Basic model

The utility of a customer of type i who consumes quantity q_i and pays t_i is given by

$$\theta_i u(q_i) - t_i,$$

where u is an increasing and strictly concave ($u' > 0$ and $u'' < 0$) of the consumption level q_i, while θ_i is a parameter reflecting the valuation of the good. When the consumer does not buy the good he has a given utility level u_0 (the same for all types) which we can normalize to zero without loss of generality. There are two categories of customers, a and b, and each category consists of two types of customers, 1 and 2, differing in their willingness to pay, with a higher willingness to pay for customer 2. Overall there are thus 4 types of individuals characterized by their valuation of the good: $\theta_{1a}, \theta_{2a}, \theta_{1b}, \theta_{2b}$, with $\theta_{1j} < \theta_{2j}$, $j = a, b$. Assume for the moment that $\theta_{2a} < \theta_{1b}$ so that the two groups do not overlap; see Figure 1. In other words the consumer with the highest valuation in group a has a lower willingness to pay than the consumer with the lowest valuation in group b. The distribution of types is common knowledge and the number of type i individuals is denoted n_i, where $i = 1a, 2a, 1b, 2b$.
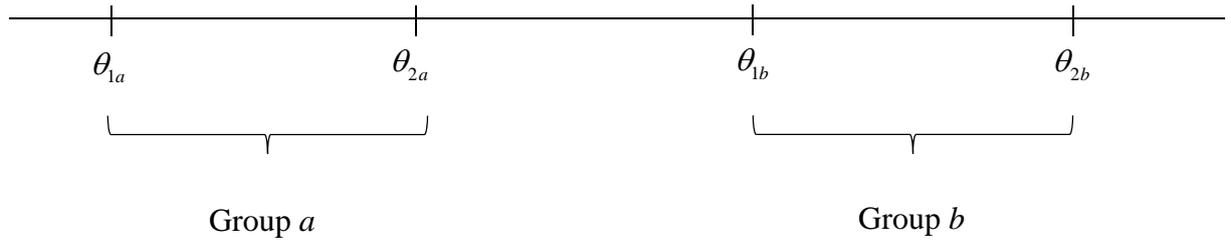
Figure 1: Separate groups

There is a single operator whose cost function is given by

$$C(q_{1a}, q_{2a}, q_{1b}, q_{2b}) = F + \sum_i n_i c_i q_i,$$

where c_i is the constant marginal cost which may or may not be the same for all types.

We consider nonlinear pricing and segmentation in the worst possible light by assuming

that the operator maximizes profit. When the operator is welfare maximizing more sophisticated pricing policies can only lead to a welfare improvement. However, when the operator is profit maximizing, there is a potential conflict between private and social objectives. A more sophisticated pricing policy then necessarily leads to a higher profit, but its impact on welfare is *a priori* ambiguous.

4. Nonlinear pricing without tagging (NLP)

With the assumption that groups are separate we have $\theta_{1a} < \theta_{2a} < \theta_{1b} < \theta_{2b}$. We assume here and throughout the paper that incentive constraints are binding according to decreasing (and adjacent) θ.[5] To simplify notation (and without loss of generality) we normalize n_i's to 1.

The profit-maximizing problem of the operator can then be written as follows

$$\max t_{1a} + t_{2a} + t_{1b} + t_{2b} - F - (c_{1a}q_{1a} + c_{2a}q_{2a} + c_{1b}q_{1b} + c_{2b}q_{2b}) \tag{1}$$

$$\text{s.t. } \theta_{2b}u(q_{2b}) - t_{2b} = \theta_{2b}u(q_{1b}) - t_{1b}, \tag{2}$$

$$\theta_{1b}u(q_{1b}) - t_{1b} = \theta_{1b}u(q_{2a}) - t_{2a}, \tag{3}$$

$$\theta_{2a}u(q_{2a}) - t_{2a} = \theta_{2a}u(q_{1a}) - t_{1a}, \tag{4}$$

$$\theta_{1a}u(q_{1a}) - t_{1a} = 0. \tag{5}$$

The decision variables are (t_i, q_i), $i = 1a, 2a, 1b, 2b$. To set up this problem we have used the standard textbook result that the participation constraint is binding for the low type (equation 5), while the utility level of the other types is determined by a series of binding incentive constraints (equations 2–4). The easiest way to solve this problem is to substitute the t_i's in the objective function from the constraints. Combining the last two constraints we obtain

$$t_{2a} = \theta_{2a}u(q_{2a}) - (\theta_{2a} - \theta_{1a})u(q_{1a}), \tag{6}$$

which together with (3) yields

$$t_{1b} = \theta_{1b}u(q_{1b}) - (\theta_{1b} - \theta_{2a})u(q_{2a}) - (\theta_{2a} - \theta_{1a})u(q_{1a}). \tag{7}$$

Combining this equation with (2) yields

$$t_{2b} = \theta_{2b}u(q_{2b}) - (\theta_{2b} - \theta_{1b})u(q_{1b}) - (\theta_{1b} - \theta_{2a})u(q_{2a}) - (\theta_{2a} - \theta_{1a})u(q_{1a}). \tag{8}$$

Substituting these expressions into the objective function (1) yields the following reformulated problem

$$\max_{q_{1a}, q_{2a}, q_{1b}, q_{2b}} \theta_{1a}u(q_{1a}) + \theta_{2a}u(q_{2a}) + \theta_{1b}u(q_{1b}) + \theta_{2b}u(q_{2b}) - F - (c_{1a}q_{1a} + c_{2a}q_{2a} + c_{1b}q_{1b} + c_{2b}q_{2b})$$

$$-(\theta_{2b} - \theta_{1b})u(q_{1b}) - 2(\theta_{1b} - \theta_{2a})u(q_{2a}) - 3(\theta_{2a} - \theta_{1a})u(q_{1a}). \tag{9}$$

The first line of this expression represents total surplus, while the second line measures the total informational rents that have to be conceded. Under full information (first degree price discrimination) where no informational rents are conceded, the operator maximizes total surplus, which it can then totally absorb. The objective function then reduces to the first line of (9). This would yield a (first-best) optimal solution which is given by

$$\theta_{2b}u'(q_{2b}^*) = c_{2b}, \tag{10}$$

[5]With two types this is necessarily the case. With more than two types this involves some assumptions on the distribution of types (similar to the monotone hazard rate assumption in the continuum case). Without such a regularity assumption we could get solutions with "bunching" (two types receive the same consumption bundles). This would make the analysis more cumbersome without affecting the results.

$$\theta_{1b} u'(q_{1b}^*) = c_{1b},\tag{11}$$

$$\theta_{2a} u'(q_{2a}^*) = c_{2a},\tag{12}$$

$$\theta_{1a} u'(q_{1a}^*) = c_{1a}.\tag{13}$$

When the customers' willingness to pay is private information the operator cannot extract the full surplus and this introduces a wedge between private and social surplus.

Differentiating (9) with respect to the consumption levels of the different types yields the following first-order conditions (FOCs)

$$\theta_{2b} u'(q_{2b}^{nt}) = c_{2b},\tag{14}$$

$$(2\theta_{1b} - \theta_{2b}) u'(q_{1b}^{nt}) = c_{1b},\tag{15}$$

$$(3\theta_{2a} - 2\theta_{1b}) u'(q_{2a}^{nt}) = c_{2a},\tag{16}$$

$$(4\theta_{1a} - 3\theta_{2a}) u'(q_{1a}^{nt}) = c_{1a},\tag{17}$$

where the superscript nt is used to refer to the non-tagged (or NLP) solution.[6]

Comparing the NLP solution defined by (14)–(17) with the first-best allocation given by (10)–(13) yields:

$$q_{2b}^{nt} = q_{2b}^*,\tag{18}$$

$$q_{1b}^{nt} < q_{1b}^*,\tag{19}$$

$$q_{2a}^{nt} < q_{2a}^*,\tag{20}$$

$$q_{1a}^{nt} < q_{1a}^*.\tag{21}$$

These expressions show that we have the first-best solution ("no distortion at the top") for the type with the highest willingness to pay, and lower than (socially) optimal consumption levels for all other types. These are all standard properties of NLP models. In particular, the downward distortion on the consumption level of any given group mitigates the informational rents of all types with higher willingness to pay.[7]

5. Nonlinear pricing with tagging (SNLP)

We now assume that the group a or b to which a customer belongs is observable and that the pricing schedule (price-quantity bundle) can be conditioned on $j = a, b$. Note that while categories a and b are observable, the position (1 or 2) of the customer inside the category is *not* observable. For example, if an individual is of type a, the operator knows that his willingness to pay parameter is either θ_{1a} or θ_{2a}, but it cannot distinguish between these two types. Observe that since the distribution of types is common knowledge, the operator also knows the distribution of willingness to pay, conditional on the group. Roughly speaking the tagging (or segmentation) thus amounts to using some observable information to get more

---

[6]The solution described by these expressions is of course only meaningful when all the q's are positive and when the consumption level increases with θ. This is where the assumptions on the distribution of θ's we mentioned above are needed. If either of these conditions is violated the solution involves bunching of two or more types. We would then have to consider a number of different cases. This does not involve any major difficulty but it would complicate the exposition and distract attention away from the main point we want to make.

[7]This can be seen from the second line of expression (9). This effect is not relevant for the type with the highest willingness to pay which explains that $q_{2b}^{nt} = q_{2b}^*$.

"precise" (albeit still incomplete) information on the customers' willingness to pay.

To determine the optimal policy in tag $j = a, b,$ we solve

$$\max_{q_{1j}, q_{2j}, t_{1j}, t_{2j}} t_{1j} + t_{2j} - F - (c_{1j}q_{1j} + c_{2j}q_{2j})$$

$$\text{s.t. } \theta_{2j}u(q_{2j}) - t_{2j} = \theta_{2j}u(q_{1j}) - t_{1j},$$

$$\theta_{1j}u(q_{1j}) - t_{1j} = 0.$$

Within each group the participation constraint of the low type and the incentive constraint of the high type are binding. Observe that we no longer have to worry about incentive constraints "between groups". For instance individuals of type $2b$ can take benefit from the pricing scheme designed for type $1b$, but they cannot benefit from any of the pricing schemes designed for group a (because the category to which the individual belongs is observable).

Combining the constraints like in the non-tagged case and substituting into the objective function we obtain in each tag $j = a, b$

$$\max_{q_{1j}, q_{2j}} \theta_{1j}u(q_{1j}) + \theta_{2j}u(q_{2j}) - F - (c_{1j}q_{1j} + c_{2j}q_{2j})$$

$$-(\theta_{2j} - \theta_{1j})u(q_{1j}).$$

Once again, the first line of this expression represents the total surplus (within the group), while the second line represents the informational rent that has to be conceded.

The FOCs are given by

$$\theta_{2j}u'(q_{2j}^t) = c_{2j} \tag{22}$$

$$(2\theta_{1j} - \theta_{2j})u'(q_{1j}^t) = c_{1j}, \tag{23}$$

where the superscript t is used to refer to the tagged (*i.e.*, SNLP) solution. Comparing this solution with the first-best allocation given by (10)–(13) yields

$$q_{2j}^t = q_{2j}^*, \tag{24}$$

$$q_{1j}^t < q_{1j}^*. \tag{25}$$

These expressions show that we now have the first-best solution for the type with the highest willingness to pay within each group, and lower than (socially) optimal consumption level for the other types.

6. Tagged vs. non-tagged solution

We are now in a position to compare the two solutions (NLP and SNLP). It is obvious that the solution with tagging yields a higher level of profit for the operator. More precisely, profits cannot be lower because the operator continues to have the option to offer the same pricing scheme as under NLP. This is of course just a fallback option for the operator and in general we can expect the solution to differ in which case profits will be strictly larger with tagging. The less straightforward question is how welfare is affected by tagging. To deal with this issue, recall that total surplus in our setting (expressed by the first line of (9)) only depends on the consumption levels of the various types. Comparing these consumption levels across solutions by making use of expressions (10)–(13), (14)–(17) and (22)–(23) yields

$$q_{2b}^{nt} = q_{2b}^t = q_{2b}^*, \tag{26}$$

$$q_{1b}^{nt} = q_{1b}^t < q_{1b}^*, \tag{27}$$

$$q_{2a}^{nt} < q_{2a}^t = q_{2a}^*, \tag{28}$$

$$q_{1a}^{nt} < q_{1a}^{t} < q_{1a}^{*}. \tag{29}$$

In words, when tagging is introduced, the solution in the highest group does not change, but we move closer to the optimal solution in the lower tag. More precisely, the high type in the low group now also consumes the efficient level, while the consumption level of the low type increases but remains below the optimal level. These results (along with the property that total surplus, S, is a concave function of any of the consumption levels) immediately implies that welfare (total surplus) is higher with tagging than without.[8] In other words, *in our setting the profit-maximizing solution under tagging (SNLP) always yields a higher level of welfare than the profit-maximizing solution without tagging (NLP).*

Intuitively, this result can be explained as follows. As shown by equations (6)–(8) the consumption level of any type, i, affects the informational rents of all the "higher" types (with a higher willingness to pay) to which type i individuals are "connected" directly or indirectly via binding incentive constraints. The optimal policy strikes a balance between surplus (which can potentially be extracted) and information rents and to reduce these rents, consumption levels are distorted downwards. From this perspective, we can easily understand that for group b nothing changes when tagging is introduced. Specifically, the q_{1b} equally affects the rents of type $2b$ in both cases (while q_{2b} does not affect any rents). Now in the untagged case q_{2a} and q_{1a} affect the rents of all types in group b and q_{1a} additionally affects the rents of type $2a$. When tagging is introduced, the link between groups is cut; inter-group mimicking is no longer possible. Consequently, q_{2a} no longer affects any rents (and is left undistorted) while q_{1a} solely influences the rents of type $2a$. This argument not only explains the result but it also indicates that it is quite robust and could easily be generalized to more general distributions of the taste parameter. In particular, it immediately follows for the case where the tags constitute a partition of a continuous distribution into two (or more intervals).

Note that this is exactly similar to age related discounts in the airline sector. Airlines can give major discounts to young passenger without worrying that this offer be used by top level executives with a high willingness to pay simply because there are few such executives in that age group.

7. Variations and extensions

To assess the robustness of this result we now consider two variations and extensions. In the first, we consider the case of overlapping groups. In the second, we assume that a simple linear tariff (the stamp price) is available to all customers and determines their reservation utility levels.

7.1 Overlapping groups

Assume now that $\theta_{1a} < \theta_{1b} < \theta_{2a} < \theta_{2b}$ so that the groups overlap in the sense that (as far as θ is concerned) the high type of the low group is above the low type of the high group; see Figure

---

[8]To see this, differentiate the first line of (9) with respect to q_i which yields

$$\frac{\partial S}{\partial q_i} = \theta_i u'(q_i) - c_i,$$

$$\frac{\partial^2 S}{\partial q_i^2} = \theta_i u''(q_i) < 0.$$

2. It is plain that this has no impact on the rules that determine tagged solution which continues to be given by expressions (22)–(23) in Section 5. Note however that the actual levels change (θ's are different so that the solution will differ). Similarly, the first-best solution is determined by the same expression as above, namely equations (10)–(13).
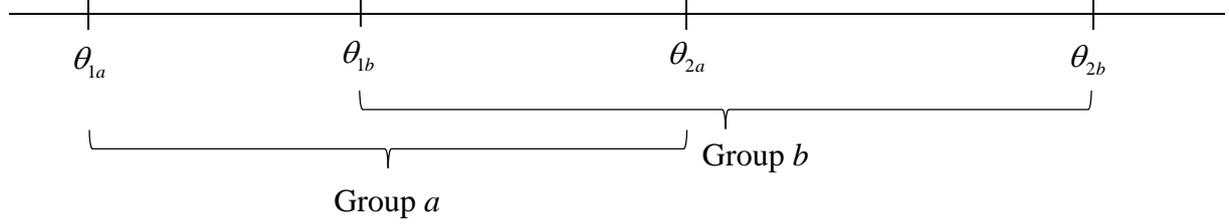

Figure 2: Overlaping groups

However, the untagged solution will change. We show in Appendix A1 that the FOCs defining the solution are now given by

$$\theta_{2b} u'(q_{2b}^{nto}) = c_{2b}, \tag{30}$$

$$(2\theta_{2a} - \theta_{2b}) u'(q_{2a}^{nto}) = c_{2a}, \tag{31}$$

$$(3\theta_{1b} - 2\theta_{2a}) u'(q_{1b}^{nto}) = c_{1b}, \tag{32}$$

$$(4\theta_{1a} - 3\theta_{1b}) u'(q_{1a}^{nto}) = c_{1a}, \tag{33}$$

where the superscript *nto* is used to denote the non-tagged (NLP) solution with overlapping groups. Comparing the different solutions we get

$$q_{2b}^{nto} = q_{2b}^{t} = q_{2b}^{*}, \tag{34}$$

$$q_{2a}^{nto} < q_{2a}^{t} = q_{2a}^{*}, \tag{35}$$

$$q_{1b}^{nto} q_{1b}^{t} < q_{1b}^{*}, \tag{36}$$

$$q_{1a}^{nto} q_{1a}^{t} < q_{1a}^{*}. \tag{37}$$

These expressions show that the three solutions yield the same level of q_{2b}, while the move to tagging increases q_{2a} and brings it closer to its optimal level. Consequently, the total surplus generated by the high type in each group increases when tagging is introduced. On the other hand, the impact of tagging on the consumption level of low types (and thus on the surplus generated by them) is ambiguous. Depending on the distribution of θ's tagging can thus increase as well as decrease consumption and total surplus of low types. To sum up, with overlapping groups the impact on welfare is ambiguous. We shall revisit this case (and show that a welfare enhancement continues to be possible) in the numerical Section 8.

7.2 Linear pricing as outside option

So far we have assumed that the reservation utility was the same for all and normalized to zero. The normalization to zero is not important and has no impact on the results. However, the assumption that the reservation utility is the same for all agents is important. In the postal sector we can think about the reservation utility as being determined by the stamp price; every agent has the option of not accepting the contract and paying the normal stamp price for its mail. The

reservation utility of an agent with parameter θ is then given by

$$v(p, \theta) = \max_q \theta u(q) - pq, \tag{38}$$

where p is the stamp price.

To determine the optimal pricing scheme (optimal contract) both without and with tagging, we now have to solve a principal-agent problem with type-dependent participation constraints which does not necessarily yield nice and simple solutions; see *e.g.*, Laffont and Martimort (2001, Section 3.3.). Depending on the profile of binding incentive and participation constraints many different regimes can occur.[9] Here, the profile of participation constraints is not arbitrary but determined by (38) so that we can obtain some results. In particular, we can look for conditions under which the various solutions considered in Sections 4 and 5 remain valid under the reformulated participation constraints. More precisely, we examine when we will obtain the same quantities for everyone while payments are shifted by a constant.

The participation constraint will be binding for individual θ_{1a} and his utility is given by $v(p, \theta_{1a})$ (rather than zero). This positive utility level adds a constant to the (profit) maximizing problem which has no impact on the FOCs and on the optimal quantities.

With the consumption profile determined in Section 4 the utility of type θ_{2a} is determined by the incentive constraint and now given by

$$(\theta_{2a} - \theta_{1a}) u(q_{1a}^{nt}) + v(p, \theta_{1a}).$$

The solution we derived continues to be valid if

$$(\theta_{2a} - \theta_{1a}) u(q_{1a}^{nt}) + v(p, \theta_{1a}) \geq v(p, \theta_{2a}). \tag{39}$$

In words, (39) says that the incentive constraint implies a utility level for θ_{2a} that is sufficiently large that the (reformulated) participation constraint is automatically satisfied. We can derive similar conditions for the other individuals as well as for the tagged solution. Note that even when the optimal quantities are not affected the sharing of the surplus is affected.

To get a better understanding of these conditions let us consider an illustration with a simple specification of preferences, namely

$$u(q) = 2\sqrt{q}. \tag{40}$$

The indirect utility is defined by

$$v(p, \theta) = \max_q \ 2\theta\sqrt{q} - pq,$$

which yields

$$q = \frac{\theta^2}{p^2},$$

and

$$v(p, \theta) = \frac{\theta^2}{p}.$$

Further, we obtain from (17) that

$$q_{1a}^{nt} = \frac{(4\theta_{1a} - 3\theta_{2a})^2}{c_{1a}^2},$$

so that condition (39) can be rewritten as

---

[9]When the reservation utility is the same for all types, a binding participation constraint for the low type along with the incentive constraints for the other types ensures that no further participation constraints are binding. When the reservation utility level depends on the type this is no longer necessarily true.

$$2(\theta_{2a} - \theta_{1a})\frac{(4\theta_{1a} - 3\theta_{2a})}{c_{1a}} + \frac{\theta_{1a}^2}{p} \geq \frac{\theta_{2a}^2}{p}.$$

Rearranging yields

$$2\frac{(4\theta_{1a} - 3\theta_{2a})}{c_{1a}} \geq \frac{\theta_{1a} + \theta_{2a}}{p}.$$

This condition will certainly be satisfied when p is sufficiently large. Note that p has to be "much" larger than c, because it includes some markup to cover the fixed cost (or part of it).

To sum up, the solutions derived above, as well as all the welfare comparisons remain valid when the linear tariff becomes available provided that the stamp price is sufficiently large.

8. Numerical example

We now present some illustrative numerical examples. They are based on the utility function

$$u(q) = \frac{q^\alpha}{\alpha},$$

with $\alpha = 1/20$.

Marginal costs are the same for all types and given by $c_{1a} = c_{2a} = c_{1b} = c_{2b} = 0.1$.

Table 1 reports the results obtained when $\theta_{1a} = 0.19 < \theta_{2a} = 0.24 < \theta_{1b} = 0.3 < \theta_{2b} = 0.41$, which corresponds to the separate group case.

	First-Best	NLP (no tag)	SNLP (tag)
q1a	1.965	0.381	1.425
q2a	2.513	1.212	2.513
q1b	3.179	1.965	1.965
q2b	4.416	4.416	4.416
CS1a		0.000	0.000
CS2a		0.953	1.018
CS1b		2.164	0.000
CS2b		4.440	2.276
Sum of CS		7.557	3.293
Profit a		7.355	7.483
Profit b		7.796	12.125
Sum of Profit		15.151	19.607
Welfare a		8.308	8.500
Welfare b		14.400	14.400
Sum of Welfare		22.708	22.901

Table 1 : Non overlapping groups - $\theta_{1a} = 0.19, \theta_{2a} = 0.24, \theta_{1b} = 0.3, \theta_{2b} = 0.41$

The first numerical column of Table 1 reports the first-best quantities, while the other two columns contain the results obtained with, respectively, the NLP (no tag) and SNLP (with tag) solutions. The numerical results (of course) confirm the analytical ranking of quantities when nonlinear prices are used; see equations (26)–(29). Furthermore, we obtain the same ranking for

welfare as for profits: tagging is better than no tagging.

We have also studied how the nonlinear optimal solutions resist to the introduction of a linear segment (the stamp price p) as an outside option in the separate group case. Intuitively, the larger the value of p, the lower th reservation utility. Starting with the solution without tagging, we investigate the minimum stamp price p such that the considered solution (NLP or SNLP) gives more utility to the four types of consumers than the linear pricing option which consists of using stamped mail. By definition, the lowest type θ_{1a} receives exactly this reservation utility level in the nonlinear solution with stamps. The minimum value of p is then $p = 0.536$ for θ_{2a}, $p = 0.331$ for θ_{1b} and $p = 0.247$ for θ_{2b}.[10] We then obtain that, even though the reservation utility (with the stamp) increases with θ, the utility (or rent) level at the nonlinear solution increases even faster with θ, so that higher θs resist better (*i.e.*, to smaller values of p) to the introduction of the stamp than lower θs.

We have performed the same exercise for the nonlinear tagged solution. Due to the tag, we only have to check two participation constraints, namely those of individuals 2 in both groups. We obtain that the minimum value of p satisfying the participation constraint with a linear pricing option is $p = 0.153$ for θ_{2a} and $p = 0.186$ for θ_{2b}. Consequently, unlike in the tagging solution, the value of p actually increases with θ. Observe that the value of p is lower, for a given θ_{2j}, with tagging than without tagging. We can see from Table 1 that the tagged solution gives more utility than the non-tagged to agents 2 in group a, allowing this solution to better resist the introduction of a linear alternative. As for group b, the utility level of θ_{2b} is actually lower with tagging than without, but the participation constraint is also different, since we add the larger $v(p, \theta_{1b})$ to the utility with tagging, compared to $v(p, \theta_{1a})$ without tagging. This latter effect is larger than the former in our setting, resulting in a (slightly) lower minimum value of p with tagging than without.

We now move to the numerical examples with overlapping groups. The easiest way to proceed is to take the same four values of θ, but to swap the values of θ_{2a} and θ_{1b}, to obtain $\theta_{1a} = 0.19 < \theta_{1b} = 0.24 < \theta_{2a} = 0.3 < \theta_{2b} = 0.41$. The results are reported in Table 2.

---

[10]To put these numbers in perspective, they are all much smaller than the profit-maximizing linear price, which equals 2.

	First-Best	NLP (no tag)	SNLP (tag)
q1a	1.965	0.381	0.791
q2a	3.179	1.965	3.179
q1b	2.513	1.212	0.687
q2b	4.416	4.416	4.416
CS1a		0.000	0.000
CS2a		2.164	2.174
CS1b		0.953	0.000
CS2b		4.440	3.337
Sum of CS		7.557	5.511
Profit a		7.428	7.542
Profit b		7.723	9.696
Sum of Profit		15.151	17.237
Welfare a		9.593	9.716
Welfare b		13.116	13.033
Sum of Welfare		22.708	22.748

Table 2 : Overlapping groups - $\theta_{1a} = 0.19, \theta_{2a} = 0.3, \theta_{1b} = 0.24, \theta_{2b} = 0.41$

The solutions without tagging are not affected by this change, except for the renaming of θ_{1b} into θ_{2a}, and vice versa. Total consumer surplus, profit and welfare are not affected by this renaming of θs, but of course their distribution among the two categories a and b is affected. As for the separate groups case, we continue to obtain a higher welfare level with tagging than without it.

We have seen in Section 1 that the solution with tagging need not always dominate the one without tagging when groups overlap. Our conjecture is that tagging dominates NLP when the overlap between the groups is "small enough", but that it may perform less well than NLP when the overlap is "large enough". To check this conjecture, we have modified the distribution of θs who now belong to the set $\{0.16, 0.21, 0.3, 0.41\}$. Table 3 reports the results with the separate case, where we know from theory that tagging dominates. As for the overlapping case, observe that the overlap is larger than in the case studied in Table 2, since it now concerns the interval $[0.21, 0.3]$ rather than $[0.24, 0.3]$ as previously. Table 4 shows that, with overlapping groups, the solution with tagging now performs less well, from a welfare viewpoint, than the solution without tagging. This confirms our conjecture that tagging is better for welfare when the division of types into groups is close to a partition with respect to the willingness to pay. In other words, segmentation (tagging) is more likely to be welfare improving when it conveys "precise" information about the unobserved distribution (that is when tags do not overlap or when overlap is not too important).

	First-Best	NLP (no tag)	SNLP (tag)
q1a	1.640	0.089	1.106
q2a	2.184	0.282	2.184
q1b	3.179	1.965	1.965
q2b	4.416	4.416	4.416
CS1a		0.000	0.000
CS2a		0.886	1.005
CS1b		2.575	0.000
CS2b		4.851	2.276
Sum of CS		8.312	3.281
Profit a		5.854	6.249
Profit b		6.974	12.125
Sum of Profit		12.828	18.374
Welfare a		6.740	7.254
Welfare b		14.400	14.400
Sum of Welfare		21.140	21.655

Table 3 : Non overlapping groups - $\theta_{1a} = 0.16, \theta_{2a} = 0.21, \theta_{1b} = 0.3, \theta_{2b} = 0.41$

	First-Best	NLP (no tag)	SNLP (tag)
q1a	1.640	0.089	0.184
q2a	3.179	1.965	3.179
q1b	2.184	0.282	0.089
q2b	4.416	4.416	4.416
CS1a		0.000	0.000
CS2a		2.575	2.573
CS1b		0.886	0.000
CS2b		4.851	3.543
Sum of CS		8.312	6.116
Profit a		6.260	6.388
Profit b		6.568	8.559
Sum of Profit		12.828	14.947
Welfare a		8.836	8.961
Welfare b		12.304	12.102
Sum of Welfare		21.140	21.063

Table 4 : Overlapping groups - $\theta_{1a} = 0.16, \theta_{2a} = 0.3, \theta_{1b} = 0.21, \theta_{2b} = 0.41$

9. Conclusion

We have studied pricing policies which combine market segmentation (tagging) with nonlinear pricing. More precisely, we have assumed that the operator can group customers into a certain number of categories on the basis of an exogenously observable characteristic at no cost. We have characterized the solution under SNLP in a four types, two categories setting and compared it to the standard NLP solution. We have shown that when the groups do not overlap (the highest willingness to pay in the low group is smaller than the lowest willingness to pay in the high group) the profit maximizing SNLP solution yields a higher welfare (total surplus) than the standard NLP solution. While we have restricted our attention to the simplest possible setting, this result could easily be generalized to more general distributions of the taste parameter (and in particular to the case of a continuous distribution). This result shows that the use of nonlinear pricing within categories may eliminate the well-known ambiguity regarding the welfare impact of third-degree price discrimination.

We then have considered two extensions. In the first, we have analyzed the case of overlapping groups. We have shown that the total surplus generated by the high type in each group increases when tagging is introduced. On the other hand, the impact of tagging on the consumption level of low types (and thus on the surplus generated by them) is ambiguous. We have provided an example where the total welfare impact continues to be positive. We have also shown that when the overlap is sufficiently significant, the opposite result can obtain. Consequently, the welfare impact of tagging in that case is essentially an empirical question which hinges on the distribution of the taste parameter and the capability of the operator to design the tags in a sufficiently informative way. When the observable characteristic is sufficiently revealing to partition the customers into separate groups according to their (unobservable) willingness to pay, segmentation is always welfare improving. In the second extension, we have assumed that a simple linear tariff (the stamp price) is available to all customers and determines their reservation utility levels. We have shown that this specific feature of the postal sector does not affect our results provided that the stamp price is sufficiently large.

To sum up, this paper shows that conceding more pricing flexibility to a profit-maximizing operator who is able to categorize its customers according to their valuation of postal services, can be welfare enhancing. In the current context of declining mail volumes resulting from the competition of electronic means of communication, the positive impact of such discriminatory pricing policy on volume should be taken into account and considered as an instrument to safeguard universal postal service. Hence, identifying relevant customer segments and determining the value of postal services for those customers become a relevant issue for postal operators. Regulators have a role to play in facilitating price-discrimination when it is welfare enhancing, by assessing pricing policies according to their economic impact rather than by adopting a dogmatic point of view based on some non-discrimination principle.

Appendix

A1. Derivation of the FOCs for the overlapping groups case

Continuing to assume that incentive constraints are binding according to decreasing θ, we now get a different pattern of binding constraints. Accordingly the profit maximizing problem of the operator is now given by

$$\max t_{1a} + t_{2a} + t_{1b} + t_{2b} - F - (c_{1a}q_{1a} + c_{2a}q_{2a} + c_{1b}q_{1b} + c_{2b}q_{2b}) \qquad (A1)$$

$$\text{s.t. } \theta_{2b}u(q_{2b}) - t_{2b} = \theta_{2b}u(q_{2a}) - t_{2a}, \qquad (A2)$$

$$\theta_{2a}u(q_{2a}) - t_{2a} = \theta_{2a}u(q_{1b}) - t_{1b}, \qquad (A3)$$

$$\theta_{1b}u(q_{1b}) - t_{1b} = \theta_{1b}u(q_{1a}) - t_{1a}, \qquad (A4)$$

$$\theta_{1a}u(q_{1a}) - t_{1a} = 0. \qquad (A5)$$

Proceeding exactly like in Section (4), namely substituting the t_i's in the objective function from the constraints and then differentiating with respect to q_i's, we obtain conditions (30)–(33)

References

Aguirre, I., S. Cowan and J. Vickers, (2010), Monopoly Price Discrimination and Demand Curvature, *American Economic Review,* **100**, 1601–15.

Cremer, H. F. Gahvari and J.M. Lozachmeur, (2010), Tagging and Income Taxation: Theory and an Application, *American Economic Journal: Economic Policy,* **2**, 31–50.

Crew, M.A. and P. Kleindorfer, (2011), Competitive strategies under FMO and intermodal competition, in *Reinventing the Postal Sector in an Electronic Age*, M. A. Crew and P. R. Kleindorfer, ed., Northampton: Edward Elgar Publishing, Inc., 1–19.

Crew, M.A. and P. Kleindorfer, (2012), Non-linear pricing, volume discounts and the USO under entry, in *Multi-modal Competition and the Future of Mail*, M. A. Crew and P. R. Kleindorfer, ed., Northampton: Edward Elgar Publishing, Inc., 1–17.

Duncan, F., D. Longman, L. Mautino and P. Dudley, (2011), Price discrimination and price differentitation within direct mail, in *Reinventing the Postal Sector in an Electronic Age*, M. A. Crew and P. R. Kleindorfer, ed., Northampton: Edward Elgar Publishing, Inc., 243–256.

Eccles, R., (2012), Price differentiation: what is acceptable for a universal service provider? in *Multi-modal Competition and the Future of Mail*, M. A. Crew and P. R. Kleindorfer, ed., Northampton: Edward Elgar Publishing, Inc., 18–31.

Geradin, D., (2011), Price discrimination in the postal sector and competition law, in *Reinventing the Postal Sector in an Electronic Age*, M. A. Crew and P. R. Kleindorfer, ed., Northampton: Edward Elgar Publishing, Inc., 230–242.

Laffont, J.-J. and D. Martimort, (2001), *The Theory of Incentives: The Principal-Agent Model*, Princeton University Press.

Willig, R.D., (1978), Pareto-superior nonlinear outlay schedules, *Bell Journal of Economics, 9,* 56–69.