DOI: 10.15826/umj.2021.1.005

SCREENING IN SPACE: RICH AND POOR CONSUMERS IN A LINEAR CITY¹

Sergey Kokovin^{\dagger}, Fedor Vasilev^{\dagger †}

HSE University,

3a Kantemirovskaya str., St. Petersburg, 194100 Russia

 $^\dagger skokov7@gmail.com, \ ^\dagger \dagger vasilevfyo@gmail.com$

Abstract: Unlike standard models of monopolistic screening (second-degree price discrimination), we consider a situation where consumers are heterogeneous not only vertically, in their willingness to pay, but also horizontally, in their tastes or "addresses" a la Hotelling's Linear City. For such a screening game, a novel model is composed. We formulate the game as an optimization program, prove the existence of equilibria, develop a method to calculate equilibria, and characterize their properties. Namely, the solution structure of the resulting menu of contracts can be either a "chain of envy" like in usual screening or a number of disconnected chains. Unlike usual screening, "almost all" consumers get positive informational rent. Importantly, the model can be extended to oligopoly screening.

Keywords: Screening, Price discrimination, Spatial competition, Linear city, Principal-Agent model, Non-convex optimization.

1. Introduction

Motivation. In economic practice, screening or "second-degree price discrimination" is quite usual in many industries. It typically generates a "product line", which is a menu of quantity-price or quality-price "packages". E.g., "packages" can mean various bottles of a soft drink offered to a heterogeneous consumers' population: (300ml for \$0.5), (450ml for \$0.8), (1000ml for \$0.95), and so on. Profit-maximizing product lines in telephony, clothes, cars, everywhere typically demonstrate some price discounts for higher quantity or quality. Why? To explain discounts and to construct product lines, economists exploit knowledge about multiple consumer types, each type being described by its "willingness-to-pay" (monetary valuation function) for higher quantity or quality. Existing types are known, but who belongs to which type is hidden from the seller; consumers self-select based on this asymmetric information — this kind of game is called "screening".

In economic theory, the standard model explaining screening or product lines dates back to Michael Spence [7]. The model is reproduced in many textbooks, monographs [5], and reviews [8]. Typically, this theory focuses on a monopolistic seller and exploits the Spence–Mirrlees "verticalordering" assumption: types of consumers are numbered in such a way that a higher number has a higher derivative of its valuation function everywhere. Then, the basic finding is the "Chain-rule theorem" about the list of active constraints in the profit-maximizing menu of packages. This list constitutes a "chain of envy" among consumers: the highest type is almost eager to switch to his/her lower neighbor's package, who in turn is almost eager to switch to his/her lower neighbor's package, and so on, other constraints being redundant. As a result, the solution method is clear

¹The authors are grateful to Pavel Ilinov, Igor Bykadorov, Mikhail Martyanov, Pavel Molchanov for discussions and help in checking the proofs. The study was financed by the HSE University Basic Research Program.

and the properties of equilibria are definite. Only the highest consumer type gets his/her Paretoefficient quantity, others are under-served. However, everybody except the lowest type gets some informational rent, i.e., a consumer surplus born by asymmetric information.

Among extensions of this theory, [4] shows that the named properties and the "chain of envy" itself are not guaranteed in situations where the standard Spence-Mirrlees assumption does not hold. Other extensions are also devoted to revealing various solution structures and properties (see [8]).

The present paper considers one more class of non-standard, poorly studied situations. Here the vertical heterogeneity of consumers (i.e., ordered valuations of types) is combined with their horizontal differentiation in some space of tastes or locations. Thus we bridge the theory of screening with another theoretical tradition, horizontal "Linear City". The latter was pioneered by [2] and includes plenty of models (for review, see [9]).

The rationale for combining both "vertical" and "horizontal" theories lies in the realism of the combination. For instance, a typical city includes consumers who are heterogeneous not only in their willingness to pay for quantity/quality but also in their geographical location. When designing his/her product lines, each seller should have in mind not only possible switches of consumers among the packages of his/her product line, but also a possibility to attract more consumers by lower prices of the whole product line. The interaction between these possibly contradicting selling strategies is the main idea here, a novel theoretical question worth studying.

Setting. Our consumers are differentiated in two ways: vertically (e.g., rich and poor) and horizontally. In a certain industry, the horizontal dimension may mean not only geography but also "tastes locations", which include some other characteristic, e.g., size of clothes. Say, T-shirts may be differentiated in qualities (vertical dimension) and sizes (horizontal dimension), suiting various consumer tastes. Still, we stick to expressive geographical interpretation, bearing in mind that tastes interpretation is isomorphic.

In the vertical dimension, we assume a finite number of consumer types. E.g., these types can be "rich", willing to pay more for quality, "middle-class", and "poor". In the horizontal dimension there is a continuum of locations among each type of consumer, uniformly distributed on the real line, the monopolistic seller being located at 0. The consumers bear some transportation costs, paying with their time and effort to go shopping. Therefore, "the farthest customer" is one whose net-of-price willingness-to-pay for the commodity almost equals his/her disutility of walking to the shop. This trade-off generates a negative dependence of the seller's range of service upon its price, interrelated with "envy" among rich and poor.

Results. We propose a novel model for such "Principal-Agents" games and formulate it as the Principal's optimization program under the Participation constraints and Incentive–Compatibility constraints (IC constraints). Such optimization (traditionally) replaces a game among agents. Further, we reduce the Principal's program to a convenient form, show why a solution should exist. Discussing a solution method, we note that our optimization problem need not be a convex one, which brings complications. In principle, one should search among all possible envy structures, which are all possible combinations of constraints. We propose a heuristic directed-search method to determine the set of active constraints in a smaller number of steps than a complete search. It exploits the first-order conditions in such a way as not to miss the global maximum.

The computational issues are the necessary preliminaries, but the economic properties of solutions are the main goal of theoretical studies in screening. For our model, the important finding is that prices tend to be lower than under screening without space. Unlike the standard setting, almost all consumers (except "the farthest customer" of each type) get some informational rent, which is the consumer surplus. Moreover, we characterize the condition on parameters that generate equilibria with Pareto-efficient sizes of packages not only for the highest type but also for some other types. In this case, the usual chain structure of envy among consumer types is broken and the graph of the solution takes the form of a broken chain, some segments being disconnected. Examples illustrate such disconnected equilibria. The highest type in each segment gets the Paretoefficient size of his/her package, whereas all other agents get distorted sizes. Why should we care about these properties? Because knowledge about Pareto-inefficiency is needed for market regulation, while knowledge about the structure of solutions can help sellers to choose their pricing policies.

Extensions and other approaches. Last but not least motivation: the present setting is one of the possible ways to develop screening in oligopoly, which remains a difficult goal for theorists to reach. In this field, the well-known paper by Rothschild and Stiglitz [6] is probably a good description of a competitive market of insurance, but it does not fit well the commodity markets because it predicts zero profit for sellers, which is not realistic. Our spatial approach appears as a good alternative because space softens competition and allows for realistic positive profits. Another approach to screening in oligopoly was recently developed in an important paper by Chade and Swinkels [1]. Their firms are heterogeneous; as a result, higher types of firms serve higher consumer types (sorting). Payoffs in such a game are not quasi-concave, but equilibria do exist. Certain firm types distort their allocations downwards; the welfare effects of private information differ from those under monopoly. The approach by Chade and Swinkels is an alternative (non-spatial) way to describe oligopoly screening with positive profit, which complements our approach.

The next section introduces the model. Subsequent sections provide examples, show that equilibria exist, provide a method to find them, and discuss equilibria properties.

2. Screening in a "Linear City": model and equilibria existence

"Linear City" in theory is typically a continuously inhabited interval, or, like here, the real line $(-\infty, \infty)$, where location 0 is the "city center". For simplicity, we restrict our attention to *uniform* distribution of each consumer type. Our single seller is located in the center, at 0. So, by symmetry, it is sufficient to formally represent only one side of this city $[0, \infty)$, the other side $(-\infty, 0]$ just mirrors the first one.²

Operating at 0, our monopolistic seller (a shop) serves the interval $[0, \infty)$ uniformly populated by each of n consumer types. Each individual is characterized by his/her type $i \in \{1, 2, ..., n\}$ and by location $\xi_i \in [0, \infty)$, which is his/her distance from the seller. The seller offers at 0 some commodity or service to all consumers, constructing a menu $\{(q_1, T_1), ..., (q_n, T_n)\}$, where each package includes quality/quantity/size q_i of the commodity and tariff or price T_i . (For instance, one may think of a shop selling a soft drink. Then, the size q_1 is destined to the "least thirsty" consumers, $q_2 > q_1$ should serve "moderate thirsty" ones, while $q_3 > q_2$ should serve "very thirsty" consumers, or very rich ones, eager to pay more.) The "transportation cost" $\tau \xi_i$ is proportional to distance (with $\tau > 0$). It is interpreted as the customers' time/money, spent traveling for shopping at point 0.³

Each point is inhabited by some mass $m_1 > 0$ of consumers type #1, by some mass $m_2 > 0$ of consumers type #2, mass $m_3 > 0$ of consumers type #3, and so on. (If one interpret m_i

²Choosing an infinite city instead of [-1, 1] means that we want to model sufficiently *high transportation costs*, to make the city edges *not served* by the seller. We are going to investigate also the case of a city completely covered by service, and the most interesting topic — two oligopolists, competing for the whole Linear City.

³As we had said, another typical interpretation of Linear City $[0, \infty)$ is some space of characteristics, e.g., sizes of clothes. In this case, the "transportation cost" means *disutility* from inappropriate size, expressed in money.

On the other side of the counter, the "Principal" in this game, the monopolistic seller, faces a marginal cost $c \ge 0$, which is his/her production cost per unit of quantity/quality. However, we shall subtract costs from valuations/tariffs and further deal only with net-of-cost valuations v and net-of-cost tariffs t^4 :

$$v_1[q] \equiv \tilde{v}_1[q] - cq, \quad \dots, \quad v_n[q] \equiv \tilde{v}_n[q] - cq[q],$$

$$t_1 \equiv T_1 - cq_1, \quad \dots, \quad t_n \equiv T_n - cq_n.$$

Respectively, the subsequent analysis looks as if costs were zero (c = 0), but the non-zero case is also included in consideration because the summand cq_i enters both sides of important constraints and cancels out.

Assumption 1 (Boundedness+). Each net valuation $v_i : R_+ \to R_+$ (i = 1, 2, ..., n) has a finite argmaximum:

$$\exists q_i^o \equiv \arg\max v_i[z] > 0.$$

Each net valuation function v_i is strictly concave, twice continuously differentiable, and it is strictly increasing on $[0, q_i^o)$, i.e., below the argmaximum.

Assumption 2 (Ordering+). The family of valuations $v_1[\cdot], v_2[\cdot], \ldots, v_n[\cdot]$ satisfies the Spence-Mirrlees ordering condition:

$$v_1'[q] < v_2'[q] < v_3'[q] \quad \forall q, \quad v_i[0] = 0 \quad \forall i = 1, 2, \dots, n.$$

As a result, graphs of all net valuations $v_i[q]$ do cross at the origin and never cross again, that is why such assumption is often called the "single-crossing condition".

Example 1. An example of valuations' family, used below for demonstrations, is a family of *affine transforms* of some common function $\nu[\cdot]$:

$$v_1[q] = a_1q + \nu[q], \quad v_2[q] = a_2q + \nu[q], \quad v_3[q] = a_3q + \nu[q],$$

with some parameters $0 < a_1 < a_2 < a_3$. E.g., it can be a family of parabolas like

$$v_1[q] = 2q - 0.5q^2$$
, $v_2[q] = 3q - 0.5q^2$, $v_3[q] = 4q - 0.5q^2$,

see similar examples below.

Traditionally for the screening theory, in such games, consumers play the role of informed "Agents", or "followers". The seller, uninformed about their types, is a "Principal", or "leader": he/she plays first, they second. He/she needs to construct a menu of "packages", being unable to discriminate among their types. Each package $(q_i, t_i) \ge 0$ includes quality q_i and tariff t_i , called also "price". One can show that there is no need to construct more packages than n agent types in the market. Non-participation is perceived as one more package (0, 0). So, the menu will consist of (0, 0) and n non-trivial elements

$$(q,t) = ((q_1,t_1),(q_2,t_2),\ldots,(q_n,t_n)).$$

⁴Hereinafter, we always use brackets like $f[\cdot]$ to denote arguments of functions, using parentheses (\cdot) for grouping.

If some packages coincide, then the consumers actually have less than n different options, but formally we shall discuss exactly n packages designed.

After the menu is set, each agent comes and "self-selects", i.e., buys a single package (q_i, t_i) from the menu in take-it-or-leave-it fashion. He/she can choose the zero outside option (not buy anything), then (0,0) brings him/her zero "reservation utility" $u_{i0} = 0$. Whenever any agent (i, ξ) chooses a package (q_i, t_i) , his/her payoff, or gross utility u_i will include his/her valuation $v_i[q_i]$ for the chosen quality q_i minus the tariff t_i , minus his/her personal transportation cost, as follows:

$$u_i[q_i, t_i, \xi] = v_i[q_i] - t_i - \tau \xi \ge u_{i0} = 0$$

Here, the parameter $\tau > 0$ is the "distance cost" coefficient for any consumer located at ξ , i.e., ξ -far from the seller located at 0. In other words, the farther is the consumer from the shop, the more he/she spends on shopping. For any type *i*, the endogenous range of service x_i is defined as the location of the farthest consumer among this type who comes to buy anything:

$$x_i: (q_i[\xi_i] > 0 \quad \forall \xi_i \le x_i), \quad (q_i[\xi_i] = 0 \quad \forall \xi_i > x_i).$$

In other words, all ξ_i located closer than x_i to 0 do buy from the seller, more distant consumers do not (the ranges of service x_i may be different among types i).

We have normalized the marginal cost c to zero (without loss of generality). So, the Principal's elementary payoff from a single purchase by consumer (i, ξ) is

$$\pi_i[q_i, t_i] \equiv m_i * (t_i - cq_i) = m_i * t_i.$$

Taking into account the endogenous range of service $x_i = x_i [q_i, t_i] \ge 0$, we are going to maximize the total Principal's expected profit, which is the integral over all consumers served (the weighted sum of individual net tariffs)

$$\Pi = \Pi[q, t, x] \equiv x_1 \cdot m_1 \cdot t_1 + \dots + x_n \cdot m_n \cdot t_n \to \max_{\{(x_i, q_i, t_i)_{i \le n} \ge 0\}}.$$
(2.1)

One may be surprised that the Principal is expected to design not only the packages but also the range of service x_i . Let us explain: traditionally for such theory, the consumers goals and behavior are expressed through inequalities. Namely, profit is maximized w.r.t. all variables simultaneously, including those chosen by consumers, under two groups of agents "rationality constraints". These are almost-standard Participation constraints (2.2) and Incentive-compatibility constraints (2.3):

$$[q_i] - t_i \ge \tau x_i \quad \forall i, \tag{2.2}$$

$$v_i[q_i] - t_i - \tau x_i \ge v_i[q_j] - t_j - \tau x_i \quad \forall i, j.$$

$$(2.3)$$

Here constraint (2.2) means that the consumer's surplus from the purchase exceeds his/her transport cost. It includes the novelty of our model: without the spatial dimension, the right-hand side of the participation constraints would be just zero. We also have in mind participation constraints

$$v_i[q_i] - t_i \ge \tau \xi_i \quad \forall \xi_i < x_i$$

for all close-to-producer agents (i, ξ) , but they are weaker than such constraint (2.2) for the farthest consumer x_i , and therefore dropped.

Any Incentive-compatibility constraint (2.3) means that a consumer i is not "envying" any other (*j*th) package, i.e., he/she has no incentive to take package j instead of one designed for him/her. The transport cost τx_i enters both sides of the Incentive Compatibility inequality, so, it can be dropped.

Now we show that the usual Chain–Rule applies here, i.e., that many Incentive-compatibility constraints can be dropped or replaced.

Lemma 1 (Chain-Rule). Solving the maximization problem (2.1)–(2.3) is equivalent to maximizing the same objective function under the following constraints:

$$v_1[q_1] - t_1 \ge \tau x_1, \quad \dots, \quad v_n[q_n] - t_n \ge \tau x_n,$$
 (2.4)

$$v_2[q_2] - t_2 \ge v_2[q_1] - t_1, \tag{2.5}$$

$$v_3[q_3] - t_3 \ge v_3[q_2] - t_2, \tag{2.6}$$

$$v_n[q_n] - t_n \ge v_n[q_{n-1}] - t_{n-1}, \tag{2.7}$$

$$q_n \ge \dots \ge q_2 \ge q_1 \ge 0, \tag{2.8}$$

where some Incentive-compatibility constraints are replaced by the ordering constraints (2.8).

P r o o f. We observe that our initial optimization problem (2.1)-(2.3) differs from the classical one only in its participation constraints. This allows us to repeat the classical proof under the Spence-Mirrlees condition (see [3]), and claim that a constraint "no-envy from any *i* to his/her lower neighbor i - 1" *implies* "no-envy" from *i* to anybody else:

$$(v_i[q_i] - t_i \ge v_i[q_{i-1}] - t_{i-1}) \Rightarrow v_i[q_i] - t_i \ge v_i[q_j] - t_j \quad \forall j$$

First, we combine constraint of "no-envy" from any i to any lower j < i with its inverse: "no-envy" from j to the higher type i:

$$v_i[q_i] - v_i[q_j] \ge t_i - t_j \ge v_j[q_i] - v_j[q_j]$$
(2.9)

and compare this with the Spence–Mirrlees condition expressed in finite differences:

$$v_i[q_i] - v_i[\tilde{q}] \ge v_j[q_i] - v_j[\tilde{q}] \quad \forall j < i, \forall (q_i, \tilde{q}) \mid q_i \ge \tilde{q}$$

We conclude that all incentive-compatible packages must satisfy the q-ordering constraint (2.8), i.e., a higher type must take a (weakly) bigger package. Thereby, adding this ordering constraint to the constraints system (2.2)–(2.3) does not influence optimization. Since our objective function is increasing in t_i and our constraints take the form $v_i[q_i] - t_i \ge \ldots$, we realize that it is sufficient to consider only intervals $q_i \le q_i^o$ below the argmaximum, where our valuations are increasing. Then, it is easy to check that bigger packages imply weakly higher tariffs for higher types: $t_1 \le t_2 \le \ldots \le t_n$ at any solution. Hence, we can ignore in optimization each "no-envy" constraint from j < i to higher *i*. Indeed, it is the right inequality in (2.9), whereas a profitable increase in both tariffs t_i, t_j can make only the left equality binding, not the right one.

Second, similarly using q-ordering and the Spence–Mirrlees condition, we check that "no-envy from any i to his/her lower neighbor i - 1" implies also "no-envy" from i to any lower type j < i - 1. Thereby, under (2.8) all non-neighboring incentive constraints are excessive, can be dropped without changing our optimization.

Thus, we have introduced a new model of screening, and represented a related Principal–Agent game as the Principal's optimization program (2.1), (2.4)-(2.8); all equilibria of our game (if any) are some profit-maximizing solutions.

3. Reduction of variables and existence of solutions

To reduce variables, one can look at the objective function (2.1) increasing in t_i , x_i , and conclude that, for each type i, the farthest-customer's participation constraint **must be active** at any

solution, i.e., become an equality (in the opposite case we could increase profit by increasing variables t_i, x_i).

So, we can use these active constraints to express the variables t_i as

$$t_1 = v_1[q_1] - \tau x_1, \quad t_2 = v_2[q_2] - \tau x_2, \quad \dots, \quad t_n = v_n[q_n] - \tau x_n.$$

Plugging these t_i into the IC constraints, we obtain their left-hand sides equal to the ranges of service x_i :

$$v_2[q_2] - t_2 = v_2[q_2] - (v_2[q_2] - \tau x_2) \equiv \tau x_2, \quad \dots, \quad v_n[q_n] - (v_n[q_n] - \tau x_n) \equiv \tau x_n.$$

Now we plug these expressions into the profit Π and into the IC constraints, thereby excluding the variables t_i and participation constraints. Thus we come to the reduced maximization problem to be solved:

$$\Pi \equiv x_{1} \cdot m_{1} \cdot (v_{1}[q_{1}] - \tau x_{1}) + \dots + x_{n} \cdot m_{n} \cdot (v_{n}[q_{n}] - \tau x_{n}) \to \max_{(x_{i},q_{i})_{i \leq n}} \text{ s.t.}$$
(3.1)

$$\tau x_{2} \geq v_{2}[q_{1}] - v_{1}[q_{1}] + \tau x_{1},$$

$$\tau x_{3} \geq v_{3}[q_{2}] - v_{2}[q_{2}] + \tau x_{2},$$

$$\dots$$

$$\tau x_{n} \geq v_{n}[q_{n-1}] - v_{n-1}[q_{n-1}] + \tau x_{n-1},$$

$$q_{n} \geq q_{n-1} \geq \dots \geq q_{1} \geq 0.$$
(3.2)

Possible solution "structures" and ideas of solving. It is common in constrained optimization to find a solution through exploring many possible combinations of constraints — inequalities, when finding out which of them will become *active* (equalities) at the true global maximum. In convex optimization, e.g., linear programming, well-known are algorithms of *directed* search among these combinations. An efficient directed search reduces the number of combinations explored, keeping a warranty of the true optimum. We are going to construct a sort of such search here. We shall denote by

$$A = \{IC_{ij}, IC_{jk}, ..., O_i, ...\}$$

any possible "solutions structure", i.e., the list (combination) of names of constraints that we assume are active at the current step. Hereinafter, IC_{ij} denotes the Incentive Compatibility constraint like $\tau x_i \geq v_i[q_j] - v_j[q_j] + \tau x_j$, and an ordering constraint like $q_i \geq q_{i-1}$ is denoted by O_i . Under any A-hypothesis, we call a related solution an "A-conditional optimum". After trying all A, we compare all such conditional optima to select a true optimum.

Unfortunately, in general, our optimization program need not be a convex one! Indeed, one can note that our objective function includes the summands $x_1 \cdot v_1[q_1]$ where both multipliers are increasing, this form precludes concavity of this function. Moreover, our constraints include the difference $v_{i+1}[q_i] - v_i[q_i]$ (of concave functions) that need not be convex or concave without an additional assumption. Generally, our domain for variables is not necessarily convex. However, our specific problem often allows for some simplifications. We start discussing them with possible empty set of active constraints $A = \emptyset$.

Disconnected kind of solutions. To introduce additional notions and notations before our existence theorem, we now show some specific, "disconnected" type of solutions (equilibria), that may occur under some specific valuations v_i .

Let us suppose that all incentive constraints (IC) and all ordering constraints $q_i \ge q_{i-1}$ are *inactive*, play no role in the solution. Then, optimization in q alone would give us the so-called

"Pareto-optimal" package sizes q_i^o , because they maximize $v_i[q_i]$ per se:

$$q_3^o \equiv \arg\max_{z \ge 0} v_3[z] \ge q_2^o \equiv \arg\max_{z \ge 0} v_2[z] \ge q_1^o \equiv \arg\max_{z \ge 0} v_1[z] > 0, \tag{3.3}$$

$$t_3^o \equiv \max_{z \ge 0} v_3[z]/2 \ge t_2^o \equiv \max_{z \ge 0} v_2[z]/2 \ge t_1^o \equiv \max_{z \ge 0} v_1[z]/2 > 0.$$
(3.4)

Using these known values q, t, now we can optimize in x_i each summand $x_i \cdot m_i \cdot (v_i[q_i^o] - \tau x_i)$ and obtain "Pareto-optimal" ranges x_i^o of service

$$x_1^o \equiv \frac{v_1[q_1^o]}{2\tau}, \quad x_2^o \equiv \frac{v_2[q_2^o]}{2\tau}, \quad \dots \quad x_n^o \equiv \frac{v_n[q_n^o]}{2\tau}.$$
 (3.5)

Such a solution means that, out of common benefit $v_i[q_i^o]$ from their contract, the closest-to-theseller consumer gets one half (as "consumer surplus"), and the seller gets the other half as his/her profit. More distant consumers get less.

To find what kind of valuations may generate such disconnected equilibria, we plug these expressions q_1^o, t_1^o into the incentive constraints. Thus we get a necessary condition (3.6) on such valuations:

$$\frac{v_{i+1}[q_{i+1}^o]}{2} - \frac{v_i[q_i^o]}{2} \ge v_{i+1}[q_i^o] - v_i[q_i^o] \Rightarrow$$

$$v_{i+1}[q_{i+1}^o] \ge 2v_{i+1}[q_i^o] - v_i[q_i^o] \qquad (3.6)$$

for all *i*. Is the inequality plausible, is a *disconnected solution possible* under any valuations v_i ? The following example confirms this.

Example 2. (Separated types #1, #2, #3.) The following example with three quadratic valuations $v_i[q_i] = a_i * q_i - b_i * q_i^2$ shows a disconnected structure:

$$a_1 = 2; \quad a_2 = 2.2; \quad a_3 = 2.3; \quad b_1 = 5; \quad b_2 = 2; \quad b_3 = 0.8.$$

The masses of types are m1 = m2 = m3 = 1, and the costs are c = 0; $\tau = 1$.

These data and direct calculations yield the following profit-maximizing Pareto-optimal sizes/tariffs:

$$q_1^* = q_1^o \equiv \arg\max_q v_2[q] = 0.2, \quad t_1 = 0.1,$$
$$q_2^* = q_2^o \equiv \arg\max_q v_2[q] = 0.55, \quad t_2 = 0.3025,$$
$$q_3^* = q_3^o \equiv \arg\max_q v_3[q] = 1.4375, \quad t_3 = 0.826563$$

Fig. 1 exhibits our Example 2 with a disconnected equilibrium. It gives also some geometry intuitions for this kind of solutions and for our optimization problem *per se*.

Thick red, green, and blue dots are the consumers' equilibrium packages: the quantity q_i lies on the horizontal axis, the tariff t_i lies on the vertical one. The valuations $v_i[\cdot]$ of the first, second, and third consumers are the solid curves painted red, green, and blue, respectively. Each dashed curve shows the equilibrium level of the valuation function for one consumer among this type, namely, for one closest to the seller. The lower is the dashed curve, the better for the consumer because its difference in height with related solid curve demonstrates the consumer's surplus (payoff).

Small red, green, and blue squares demonstrate the equilibrium packages which would occur under standard, space-less screening. Comparing standard and new outcomes, we observe that

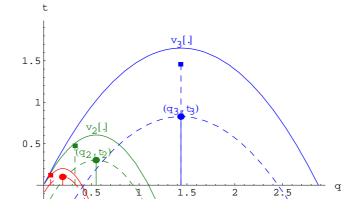


Figure 1. Valuations and solutions.

adding space to screening can diminish distortion. Namely, space increases quantities q and also brings benefits for the close-to-seller consumers by diminishing tariffs t.

Example 3. (Connected types #1+#2, separated #3.) This example is almost the same as Example 2, only $b_2 = 3$. The quadratic valuations are

$$v_i[q_i] = a_i * q_i - b_i * q_i^2$$

with

$$a_1 = 2;$$
 $a_2 = 2.2;$ $a_3 = 2.3;$ $b_1 = 5;$ $b_2 = 3;$ $b_3 = 0.8,$

the same masses m1 = m2 = m3 = 1 and costs c = 0; $\tau = 1$. The solution shows a partially disconnected structure:

$$q_1^* = 0.19 < \arg\max_q v_2[q], \quad t_1 = 0.1104336,$$

$$q_2^* = q_2^o \equiv \arg\max_q v_2[q] = 0.366667, \quad t_2 = 0.19712,$$

$$q_3^* = q_3^o \equiv \arg\max_q v_3[q] = 1.4375, \quad t_3 = 0.826563.$$

Example 4. (Connected types #1+#2+#3.) This example is almost the same as Example 3, only $b_3 = 2$. The quadratic valuations are

$$v_i[q_i] = a_i * q_i - b_i * q_i^2$$

with

$$a_1 = 2; \quad a_2 = 2.2; \quad a_3 = 2.3; \quad b_1 = 5; \quad b_2 = 3; \quad b_3 = 2,$$

the same masses m1 = m2 = m3 = 1 and costs c = 0; $\tau = 1$. It shows a completely connected solution structure:

$$q_1^* = 0.182261 < \arg\max_q v_2[q], \quad t_1 = 0.10786,$$

$$q_2^* = 0.34472 < \arg\max_q v_2[q], \quad t_2 = 0.208432,$$

$$q_3^* = q_3^o \equiv \arg\max_q v_3[q] = 0.575, \quad t_3 = 0.31449.$$

Economic intuitions. Economically, why the "chain of envy", quite usual in standard screening, can be broken in spatial screening? Why disconnected profit-maximizing solutions may happen? Well, we know that in usual screening there are no reasons for the principal to leave any incentive compatibility constraint, or the 1st participation constraint, inactive. Such a plan would bring waste of profit. Respectively, the 1st type in usual screening gets zero informational rent, nothing beyond the reservation utility.

By contrast, in our spatial screening model, all closer-than-the-farthest customers get equilibrium payoffs *higher* than their reservation ones; i.e., the tariff for them is lower than it could be. This slack is not wasted, from the view of the profit-maximizing Principal, it is a sacrifice for extending his/her service range, his/her coverage of consumers. This objective, *constructing a utility slack to attract more consumers*, can make one or more incentive constraints inactive. The realism of such a trade-off in many markets — is the main reason for building our new model of screening.

Our discussion of examples and reasons for non-active constraints ends up with the following conclusion. Generally, the list A of active constraints may be empty, or include all IC constraints $(A = \{IC_{n,n-1}, IC_{n,n-1}, ..., IC_{21}\})$, or may consist of various combinations of active constraints.

Equilibria existence. Returning to the general case with unknown A, we should ask: are there always solutions $\{(x_i, q_i)_{i \leq n}\}$ to our maximization program (3.1)–(3.2)? Our domain is not empty. Indeed, to show a sample admissible plan we can take all partial Pareto-optimal sizes q_i^o . These "separate maxima" exist under our assumptions, they satisfy our ordering.⁵ We can supplement these q with sufficiently small x like $x_1 = x_2 = x_3 = \varepsilon < \min_j v_j [q_j^o]/\tau$. So, a positive admissible plan (q, x) > 0 exists. Moreover, it brings a strictly positive profit $\Pi > 0$.

Now studying our objective function, we observe that the highest-type quality q_n enters Π only once, so, this variable must take its Pareto-optimal value $q_n = q_n^o$ at any solution. As to the other arguments q_i of the objective function, we conclude that we can restrict them as $q_i \leq q_i^o$ without sacrificing our objective function (since, for higher values of q, our profit Π becomes smaller). Similarly, without sacrificing our objectives, variables x can be restricted as $(v_i[q_i] - \tau x_i) \geq 0$, i.e., $x_i \leq v_i[q_i]/\tau \leq v_i[q_i^o]/\tau$.

Therefore, without loss of optima, we can squeeze our initial domain (positive orthant), and maximize now our function $\Pi(q, x)$ on a restricted compact domain K constructed as

$$K \equiv \{ (q, x) | 0 \le q_i \le q_i^o + 1, \quad 0 \le x_i \le x_{max} \equiv \max_i v_j [q_j^o] / \tau + 1 \quad \forall i = 1, 2, 3 \}$$

Our objective function $\Pi(q, x)$ is continuous. Therefore, by Weierstrass's extreme value theorem, we have established the following statement.

Proposition 1. Under our assumptions, there exists a solution to our maximization problem (3.1)-(3.2).

Moreover, one can conclude from our discussion that the maximum lies strictly below the artificial upper bounds $q_i^o + 1$, x_{max} (being an inner solution) and brings a positive profit Π .

How can we practically find a solution under any specific valuation functions v_i ?

Of course, one can exploit any iterative or exact numerical method. Since we deal here with differentiable functions, it is possible to use exact finite methods, exploiting the first-order conditions, even without being sure in convex optimization. Indeed, after finding all stationary points and border solutions, one can compare (a finite number of) related local maxima, to choose the global one.

⁵We also can take sufficiently small values of $x_1 = x_2 = x_2 = \varepsilon < v_1[q_1^o]/\tau < v_2[q_2^o]/\tau < v_3[q_3^o]/\tau > 0$, and assure that the objective function *can be positive* under some values of the optimizers.

To implement this idea, solving first-order equations for *all* possible A-structures should be sufficient to find the true optimum, but this way can be computationally tedious. We shall suggest a heuristic method that explores *some* A-structures. To construct each step of this method, the next section suggests a convenient way of using First-Order Conditions (FOC) for exploring any hypothesis A about active constraints.

4. Using FOC for any hypothetical list A of active constraints

This section explains how to further reduce our variables and use First-Order Conditions to find any A-conditional maximum under some hypothesis A. We explain it by an example.

Suppose that we have n = 4 consumer types and assume that family A of active IC constraints (A explored on some step of our general search algorithm) connects three adjacent agent types $\{\#1, \#2, \#3\}$ as $A = \{IC_{21}, IC_{32}\}$, whereas type #4 is separated and ordering constraints are not binding. We find related A-conditional maximum as follows.

As we have ensured, the highest variable q_3^* among the chain $\{\#1, \#2, \#3\}$ must take its v_3 -maximizing Pareto-value q_3^o (here and further accent * denotes solutions):

$$q_3^* = q_3^o \equiv \arg\max v_3[q_3]$$

Similarly, the isolated variable q_4 also takes its Pareto-optimal value $q_4^* = q_4^o$. (By contrast, lower variables q_i in the chain need not become Pareto-optimal because the incentive constraints may be active.)

Whenever any IC_{ji} is active, we can define the difference function $V_{ji}[q_i] \equiv v_j[q_i] - v_i[q_i]$. E.g., for two IC included in $A = \{IC_{21}, IC_{32}\}$, these difference functions are

$$V_{32}[q_2] \equiv v_3[q_2] - v_2[q_2], \quad V_{21}[q_2] \equiv v_2[q_1] - v_1[q_1].$$

(Special linear case. Such difference function $V_{32}[\cdot]$ can appear *linear* in the particular case when the valuations family $v_i[\cdot]$ is built from a common function $u[\cdot]$ as its linear modification $v_i[x] = a_i x - u[x]$ with $a_1 < a_2 < \dots$ In this case, $V_{32}[q_2] \equiv a_3q_2 - a_2q_2 = (a_3 - a_2)q_2$, $V_{21}[q_1] \equiv a_2q_1 - a_1q_1 = (a_2 - a_1)q_2$).

We can invert any difference function V_{ij} because it is *increasing*, by Spence–Mirrlees assumption. We denote the inverse $\Lambda_{ij}[\cdot] \equiv V_{ij}^{-1}[\cdot]$. Further, to reformulate active constraints — equations $\tau x_3 = v_3[q_2] - v_2[q_2] + \tau x_2$ through these functions Λ_{ij} , we express the volumes q_i through the differences in service ranges:

$$q_{2} = \Lambda_{32} \left[\tau x_{3} - \tau x_{2} \right] \equiv V_{32}^{-1} \left[\tau x_{3} - \tau x_{2} \right],$$

$$q_{1} = \Lambda_{21} \left[\tau x_{2} - \tau x_{1} \right] \equiv V_{21}^{-1} \left[\tau x_{2} - \tau x_{1} \right].$$

Using this transform to simplify our optimization, we can get rid of all variables except the service ranges x_i :

$$\tau \Pi = \tau x_1 \cdot m_1 \cdot \left(v_1 [\Lambda_{21} [\tau x_2 - \tau x_1]] - \tau x_1 \right) + \tau x_2 \cdot m_2 \cdot \left(v_2 [\Lambda_{32} [\tau x_3 - \tau x_2]] - \tau x_2 \right) + \tau x_3 \cdot m_3 \cdot \left(v_3 [q_3^{opt}] - \tau x_3 \right) \to \max_{\substack{x = (x_1, x_2, x_3) > 0}},$$

$$(4.1)$$

and deal with unconstrained optimization. We check after finding the unconstrained maxima whether the ordering conditions and out-of-A IC constraints are satisfied. In the opposite case (violated outside constraints), we reject the hypothesis A and explore another one.

In formulation (4.1), we have multiplied our objective function by τ to prepare subsequent usage of auxiliary variables $y_i \equiv \tau x_i$, δ_{ij} , to simplify the analysis. This trick explains also the following remark.

Remark 1. The parameter τ only multiplies our payoff but does not influence the main variables — maximizers (x, q, t).

Solution for a connected component $\{\#i, \#i+1, \#i+2\}$ under $A = \{IC_{21}, IC_{32}\}$. To find the solution components (x_1, x_2, x_3) , it is sufficient to differentiate the partial objective function (4.1) by (x_1, x_2, x_3) , explore these first-order conditions, and compare all resulting stationary points (the points can be multiple if the function is non-concave, but their number is finite), to choose the true maximum. Afterwards we derive the remaining variables (q, t) from these (x^*) . This gives the part $(q_i^*, t_i^*, x_i^*)_{i\leq 3}$ of the needed solution. Turning to the remaining type #4, the part $(q_i^*, t_i^*, x_i^*)_{i\leq 3}$ is supplemented with the Pareto-optimal values $(q_4^*, t_4^*, x_4^*) = (q_4^o, t_0^o, x_4^o)$ found for any isolated type from equations (3.3)–(3.5). Now we should check if the ordering conditions and the unused IC constraints (IC_{43}) are really satisfied, inactive. If it is wrong, this hypothesis $A = \{IC_{21}, IC_{32}\}$ is rejected, otherwise, it can be compared with other hypotheses.

To simplify using the first-order conditions, the objective function can be expressed in new auxiliary variables $\delta_{ji} \equiv y_j - y_i$ as:

$$\tau \Pi = (y_3 - \delta_{32} - \delta_{21}) \cdot m_1 \cdot v_1 [\Lambda_{21} [\delta_{21}]] - m_1 \cdot (y_3 - \delta_{32} - \delta_{21})^2 + (y_3 - \delta_{32}) \cdot m_2 \cdot v_2 [\Lambda_{32} [\delta_{32}]] - m_2 \cdot (y_3 - \delta_{32})^2 + (y_3) \cdot m_3 \cdot v_3 [q_3^{opt}] - m_3 \cdot y_3^2 \to \max_{\substack{y = (\delta_{21}, \delta_{32}, y_3) \ge 0}}.$$
(4.2)

If we treat the variable y_3 parametrically, it is easy to observe that the *concavity* in the remaining variables δ_{ij} of any summand is guaranteed when every function $v_i [\Lambda_{i+1,i} [z]]$ is concave for all z. Concavity may help in practical optimization as well as the following technical lemma.

Lemma 2 (Existence of solutions for a component). Given a family A of active IC constraints and its connected component $\{\#i, \#i+1, \#i+2\}$, the first-order conditions for related function formulated as (4.1) must give at least one solution that is a global maximum of this function.

Proof. Though Proposition 1 states that maxima do exist when the problem is expressed in terms of variables (q, x), the above lemma is more specific. Looking at the objective function (4.2) we note that values of $v_i [\Lambda_{i+1,i}[z]]$ are bounded from above by the maximal value $\max_{z\geq 0} v_i[z]$, whereas other terms are quadratic with minus and take arbitrarily low (negative) values when any variable approaches infinity. So, in spite of the unknown concavity of this objective function, we are sure that all its local maxima must be inner ones, not go to infinity. \Box

As to the *uniqueness* of a solution for any connected component, probably, it can be proved using the assumption of "special linear case", i.e., $\Lambda_{21}'' = 0$, but this question remains unclear.

Now we explain how to use our A-conditional solutions and connected components to sequentially search among A-conditional maxima for finding a true maximum. Related heuristic computational procedure hopefully economizes calculations. The next section also describes all possible solution structures.

5. Non-active IC constraints and method of search among broken chains

We have shown how to find conditional optima for each connected component belonging to any hypothetical family A of active constraints. Now let us show how to go step by step from one hypothetical family A to another, revealing which constraints *should be active* at the solution, i.e., building a sequence of families that approaches the optimal family A^* .

To begin with, we can try using Pareto-optimal sizes $q_i^* = q_i^o \equiv \arg \max_q v_i[q]$ with related profit-maximizing tariffs $t_i^* = t_i^o \equiv v_i[q_i^o]$ and ranges $x_i^* = x_i^o = v_i[q_i^o]/2$ as in (3.3)–(3.5). It can happen that no IC or ordering constraints are violated. Then we can declare that it is a global optimum with a *disconnected solution structure*: $A = \emptyset$, like our Example 2 (because adding any constraint to A cannot enhance the objective function).

More typically, some constraints are violated at $A = \emptyset$, then some packages should become connected, and finding the optimal solution structure A^* becomes more difficult. These considerations give us intuition for the following algorithm for finding A^* , through checking active IC constraint.

The idea of general optimization algorithm.

Let us denote by $ICO \equiv \{IC_{21}, IC_{32}, ..., IC_{n-1,n}, O_1, ..., O_n\}$ the list of all possible Incentivecompatibility and Ordering constraints in our reduced program (3.1). Now we describe an algorithm of directed search among multiple possible combinations, various families $A \subset ICO$ of *active* constraints. Each family A may generate its own A-conditional-optimal solution (q^A, x^A) , i.e., the solution under these constraints only. If it happens that this solution (q^A, x^A) does not violate other, non-included constraints $ICO \setminus A$, then we have reached an *admissible A-conditionallyoptimal* solution (local maximum). Otherwise, we reject the family A as a possible generator of solutions. After we explore ALL *admissible A-conditionally-optimal* solutions for all possible $A \subset ICO$ (exploring finite number of combinations), we can compare their profits and choose the best local maximum, being sure that it is a global maximum.

Computationally, it appears a tedious, long search. However, luckily, our specific optimization problem allows for shorter, sequential, directed search among all possible families A, starting with an empty set $A = \emptyset$ and then adding the active constraints one-by-one, going from lower to higher consumer types, as follows. We provide reasons why during this search we cannot miss the optimal system A^* of active constraints.

Heuristic algorithm of sequential search among possible A-structures.

Let us denote packages by $w_i = (q_i, t_i)$.

1. We start from the lowest type #1. To assign his/her package w_1 , we first assume that his/her upper IC_{21} is inactive (separated #1) and therefore assign the related Pareto-optimal values

$$q_1 = q_1^0, \quad t_1 = v_1[q_1^0]/2, \quad x_1 = \frac{v_1[q_1^0]}{2\tau}.$$

2. Similarly, we find the second Pareto-optimal package $w_2 = (q_2, t_2)$, assuming that both IC_{21} and IC_{32} are inactive:

$$q_2 = q_2^0, \quad t_2 = v_2[q_2^0]/2, \quad x_2 = \frac{v_2[q_2^0]}{2\tau}$$

3. Now we check whether these two packages w_1 and w_2 violate the constraint IC_{21} .

If IC_{21} is violated, then type 1 and type 2 are "connected", i.e., their packages w_1, w_2 should be optimized together, within one problem in the way explained in Section 4. In this case, we solve program (4.2) that includes these two types: $A = \{IC_{21}\}$. We already know that among these agents the highest size q_2 necessarily becomes Pareto-optimal: $q_2^* = q_2^o$, but the tariff t_2 will differ from Pareto-optimal $v_2[q_2^0]/2$. Anyway, we get some partial plan — an admissible couple (w_1, w_2) .

4. Now we find the third-type optimal package under the assumptions that IC_{32} and IC_{43} are inactive and #3 is separated. Thereby, w_3 should be Pareto-optimal:

$$q_3 = q_3^0, \quad t_3 = v_3[q_3^0], \quad x_3 = v_3[q_3^0]/2.$$

Now we check (violated or not) the constraint IC_{32} , using the packages w_3 and w_2 already found previously. If IC_{32} is violated, then type 2 and type 3 become connected. They become parts of a unified optimization problem with $A = \{IC_{32}, IC_{21}\}$ in the case if 1 and 2 were connected. At this stage, we apply program (4.2) and a related known method to the family $A = \{IC_{32}, IC_{21}\}$ and find three connected packages (w_1, w_2, w_3) . Thereby, our previously found (w_1, w_2) will change.

In the opposite case, if agents #1, #2 were not connected (separated #1), we apply program (4.2) and the related method with the smaller family $A = \{IC_{32}\}$ to find connected packages (w_2, w_3) . Then we check (violated or not) IC_{21} : if it is not violated, then the previously found Pareto-optimal w_1 does not change, otherwise, it changes. In the latter case, we again must solve the three-package component $A = \{IC_{32}, IC_{21}\}$. It means that finding w_3 may work in such a way that previously disconnected packages #1 and #2 become connected.

We argue that adding a higher component w_k to the previous locally-optimal partial plan $(w_1, ..., w_{k-1})$ may only increase connectedness, but not break it!

Anyway, the calculations above produce some partial plan — an admissible triple (w_1, w_2, w_3) for three lowest components of the desired solution $(w_1, ..., w_n)$.

- 5. Further, we proceed in the same way adding agent type #4 (package w_4) to our analysis and checking, whether #4 becomes connected with previous packages, or not. In the latter case, the previous packages remain unchanged, otherwise, they change. When they change, any disconnected (excluded) IC constraints below #4 may become connected into the solution structure.
- 6. We repeat adding new, higher agent types one-by-one and adjusting the current plan w accordingly until we reach the highest type #n. At each step, adding a higher component w_k to the previous locally-optimal partial plan $(w_1, ..., w_{k-1})$ may only increase connectedness, but not break it. When the connectedness increases, we must recalculate the lower components, otherwise, this is not necessary.

In the end of this algorithm, various outcomes are possible: all types become separated; type 1 and type 2 are linked but other separated; type #1 is separated, types #2 and #3 are linked but other separated; all types can be connected, and so on.

This algorithm gives an exact solution through a finite number of steps, each step solving equations, which are the specific first-order conditions for the related A-structure.

Commenting on the general idea, we observe that starting from the lowest types, we check the connectivity of types. When we meet a new active IC constraint, we check if any previously inactive constraint becomes active. The "impulse" of restrictions goes down through the chain of lower connected types. Indeed, when IC_{i+1} becomes active, the lower-neighbor utility $u_i[w_i]$ decreases, therefore the lower IC_i might force $u_{i-1}[w_{i-1}]$ to decrease also, and so on — this is what we mean by the "impulse".

Why should the algorithm attain the optimal list of constraints A^* ? At each step of the process, we maximize the profit (from the partial plan) with the *minimal possible* number of constraints. We always keep the partial plan admissible, always checking if any additional constraints are activated. These considerations are not complete proof, but they support the idea that the solution found by the algorithm should be the global maximum. We suppose that this algorithm reaches the true optimal solution structure A^* and the optimal plan w^* , and that we need not explore any other structures A avoided by this method. We cannot provide more detailed proof of this fact so far.

Conclusion

To summarize, this paper suggests a new model of screening, which is second-degree price discrimination, for situations where consumers are both vertically and horizontally heterogeneous; in their willingness to pay for quality and in their locations in geographical space or space of preferred characteristics of the commodity.

The screening game is reformulated as an optimization program of the seller. The existence of solutions, which are equilibria, is established under typical for the screening literature assumptions. This constrained program being potentially non-convex, the heuristic algorithm is proposed to reduce the search among all possible combinations of active constraints, "solution structures".

The examples show that solution structures can vary: agent types can be connected by one common chain of IC constraints (called "chain of envy" in the screening literature), or completely separated, or consist of several chains of adjacently-numbered agents.

The important economic feature of such equilibria in spatial screening is that, like in the usual screening, the highest (in each chain) agent type always gets a Pareto-efficient quality, whereas others do not, their quality is distorted downwards. Unlike the usual screening, almost all agents (all except "the farthest consumer") get some informational rent, their payoffs are higher than their reservation utility.

The most interesting economic extension of this study would be an application of our spatial screening model to oligopoly screening. It promises an explanation of many real-life situations in competition, poorly studied so far.

REFERENCES

- Chade H., Swinkels J. Screening in vertical oligopolies. *Econometrica*, 2021. Vol. 89, No. 3. P. 1265–1311. DOI: 10.3982/ECTA17016
- Hotelling H. Stability in competition. *Economic J.*, 1929. Vol. 39, No. 153. P. 41–57. DOI: 10.2307/2224214
- Katz M. L. Nonuniform pricing with unobservable numbers of purchases. Rev. Econ. Stud., 1984. Vol. 51, No. 3. P. 461–470. DOI: 10.2307/2297434
- Kokovin S. G., Nahata B. Method of digraphs for multi-dimensional screening. Ann. Oper. Res., 2017. Vol. 253, No. 1. P. 431-451. DOI: 10.1007/s10479-016-2320-3
- Laffont J-J., Martimort D. The Theory of Incentives: The Principal-Agent Model. Princeton: Princeton Univ. Press, 2002. 421 p. DOI: 10.2307/j.ctv7h0rwr
- Rothschild M., Stiglitz J. Equilibrium in competitive insurance markets: an essay on the economics of imperfect information. Quart. J. Econ., 1976. Vol. 90, No. 4. P. 629–649. DOI: 10.2307/1885326
- 7. Spence A. M. Job market signaling. Quart. J. Econ., 1973. Vol. 87, No. 3. P. 355–374. DOI: 10.2307/1882010
- Stole L. A. Price discrimination and competition. *Handbook Ind. Organiz.*, 2007. Vol. 3. P. 2221–2299. DOI: 10.1016/S1573-448X(06)03034-2
- Torbenko A. Linear city models: overview and typology. J. New Econ. Assoc., 2015. Vol. 25, No. 1. P. 12–38. (in Russian) URL: https://ideas.repec.org/a/nea/journl/y2015i25p12-38.html