CHAPTER 5

The Economics of Multidimensional Screening

Jean-Charles Rochet and Lars A. Stole

1. MOTIVATION AND INTRODUCTION

Since the late 1970s, the theory of optimal screening contracts has received considerable attention. The analysis has been usefully applied to such topics as optimal taxation, public good provision, nonlinear pricing, imperfect competition in differentiated industries, regulation with information asymmetries, government procurement, and auctions, to name a few prominent examples.[1] The majority of these applications have made the assumption that preferences can be ordered by a single dimension of private information, largely to facilitate finding the optimal solution of the design problem. However, in most cases that we can think of, a multidimensional preference parameterization seems critical to capturing the basic economics of the environment. For example, consider the case of duopolists in a market where each firm competes with nonlinear pricing over its product line. In many examples of nonlinear pricing (e.g., Mussa and Rosen 1978 and Maskin and Riley 1984), it is natural to think of consumers' preferences being ordered by the willingness to pay for additional units of quantity or quality. But, if we believe that competition between duopolists is imperfect in the horizontal dimension as suggested, for example, by models such as Hotelling's (1929), then we need to introduce a form of horizontal heterogeneity as well. As a consequence, a minimally accurate model of imperfect competition between duopolists suggests including *two* dimensions of heterogeneity – vertical and horizontal.

There are several additional economic applications that naturally lend themselves to multidimensional heterogeneity.

- General models of pricing. In some instances, a firm may offer a single product over which the preferences of the consumer may depend

[1] Among the seminal contributions, we can cite Mirrlees (1971, 1976) for optimal taxation, Green and Laffont (1977) for public good provision, Spence (1980) and Goldman, Leland, and Sibley (1984) for nonlinear pricing, Mussa and Rosen (1978) for imperfect competition in differentiated industries, Baron and Myerson (1982), Baron and Besanko (1984), McAfee and McMillan (1987), and Laffont and Tirole (1986, 1993) for regulation, and Myerson (1981) for auctions.

importantly on several dimensions of uncertainty (e.g., tastes, marginal utility of income, etc.). In other instances, a firm may be selling an array of distinct products, of which consumers may desire any subset of the total bundle of goods. In this latter case, the dimension of heterogeneity of consumers' preferences for the firm's products will be at least as large as the number of distinct products.

- Regulation under richer asymmetries of information. As noted in the seminal article by Baron and Myerson (1982) on regulation under private information, at least two dimensions of private cost information naturally arise – fixed and marginal costs. Another example is studied by Lewis and Sappington (1988) in which the regulator is simultaneously uncertain about cost and demand. If we wish to take the normative consequences of asymmetric information models of regulation seriously, we should check the robustness of the results to such reasonable bidimensional private information.
- Income effects and related phenomena. Many times it makes sense to think of two-dimensional information when privately known budget constraints or other forms of limited liability are present. For example, how should a seller design a price schedule when customers have random valuations and simultaneously random budget constraints?
- Auctions. Similar to the aforementioned problem, we may suppose that multiple buyers bid for a single item, but their preferences depend on a privately known budget constraint in addition to a private valuation for the good (as in Che and Gale, 1998, 1999, 2000). Or in another important auction setting, suppose (as in Jehiel, Moldovanu, and Stacchetti 1999) that a buyer's preferences depend not only on his own valuation of the good, but also on the privately known externality from someone else getting the good instead (e.g., two downstream firms bid for an exclusive franchise and the loser must compete against the winner with an inferior product). Although in this paper, we do not consider the auction literature in depth, the techniques of optimal contract design in multidimensional environments are clearly relevant.[2]

Unfortunately, the techniques for confronting multidimensional settings are far less straightforward as in the one-dimensional paradigm. This difficulty has meant that the bulk of applied theory papers in the self-selection literature are based on one-dimensional models of heterogeneity. As a consequence, the results of these economic applications remain uncomfortably restrictive and possibly inaccurate (or at least nonrobust) in their conclusions. In this sense, we have been searching under the proverbial street lamp, looking for our lost keys, not because that is where we believe them to lie, but because it is apparently the only place where we can see. This survey is an attempt to catalog and

[2] Other multidimensional auctions problems are studied by Gal, Landsberger, and Nemirovski (1999) and Zheng (2000).

explain the terrain that has been discovered in the brief forays away from the one-dimensional street lamp – indicating both what we have learned and how light or dark the night sky actually is.

In Section 2, we review the one-dimensional paradigm, emphasizing those aspects that will generate problems as we extend the analysis to multiple dimensions. In Section 3, the general multidimensional paradigm is explained for both the discrete and continuous settings. We illustrate the concepts in a simple two-type "multidimensional" model, explaining how the multidimensionality of types introduces new economic and mathematical aspects of the screening problem. In Sections 4–9, we specialize our discussion to specific classes of multidimensional models that have proven successful in the applied literature. Section 4 presents results on separation and aggregation that greatly simplify multidimensional screening. Section 5 considers environments in which there is a single, nonmonetary contracting variable, but multiple dimensions of type – a scenario that also frequently gives rise to explicit solutions. Section 6 looks at a further specialized subset of models (from Section 5) that are economically important and mathematically tractable: bidimensional private information settings in which one dimension of information enters the agent's utility function additively. Section 7 considers a series of multidimensional models that have been successfully applied to competitive environments. Section 8 considers a distinct set of multidimensional environments in which information is revealed over time. Finally, Section 9 considers the more subtle problems inherent in general models of multiple instruments and multidimensional preferences; here, most papers written to date have considered the scenario of multiproduct monopoly bundling, so we study this model in some detail. Section 10 concludes.

2. A REVIEW OF THE ONE-DIMENSIONAL PREFERENCE MODEL

Although it is often recognized that agents typically have several characteristics and that principals typically have several instruments, the screening problem has most of the time been examined under the assumption of a single characteristic and a single instrument (in addition to monetary transfers). In this case, several qualitative results can be obtained with some generality:

1. When the single-crossing condition is satisfied, only local (first- and second-order) incentive compatibility constraints can be binding.
2. In most problems, the second-order (local) incentive compatibility constraints can be ignored, provided that the distribution of types is not too irregular.
3. If bunching is ruled out, then the principal's optimal mechanism is found in two steps:
   (a) First, compute the minimum expected rent of the agent as a function of the allocation of (nonmonetary) goods.

(b) Second, find the allocation of goods that maximizes the surplus of the principal, net of the expected rent computed in (a).

To understand the difficulties inherent in designing optimal screening contracts when preferences are multidimensional, it is useful to first review this basic one-dimensional paradigm. This will serve both as a building block for the multidimensional extensions and as an illustration of how one-dimensional preferences generate simplicity and recursion in the optimization program.

We will use a simple nonlinear pricing framework similar to Mussa and Rosen (1978) as our basic screening environment, elaborating as appropriate. Suppose that a monopolist sells its products using a nonlinear tariff, $P(q)$, where q is the amount of quantity chosen by the consumer and $P(q)$ is the associated price. The population of potential consumers of the firm's good have preferences that can be indexed by a single-dimensional parameter, $\theta \in \Theta \equiv [\underline{\theta}, \overline{\theta}]$, and is distributed in the population according to the absolutely continuous distribution function $F(\theta)$, where $f(\theta) \equiv F'(\theta)$ represents the associated density. Let each consumer's preferences for consuming $q \in \mathcal{Q} \equiv [0, \bar{q}]$ for a price of P be given by

$$u = v(q, \theta) - P.$$

Note that preferences are linear in money. To place some additional structure on the effect of θ, we assume the well-known, single-crossing property that $v_{q\theta}$ has a constant sign; in this paper, we will associate higher types with higher marginal valuations of consumption; hence, $v_{q\theta} > 0$. This condition makes the one-dimensional assumption restrictive.[3] It is worth noting that this condition has two equivalent implications: (i) the indifference curves of any two types of consumers cross at most once in price-quantity space, and (ii) the associated demand curves do not intersect and are completely ordered as a family of curves given by $p = v_q(q, \theta)$. We will begin our focus on the even simpler linear-quadratic setting in which $v(q, \theta) = \theta q - \frac{1}{2}q^2$. In this case, the associated demand curves are parallel lines, $p = \theta - q$.

There are two methodologies used to solve one-dimensional screening problems – what we refer to as the parametric-utility approach and the demand-profile approach. The former has been more commonly used in the applied theory literature, but the latter provides useful conceptual insights, particularly in the multidimensional context, that are easily overlooked in the former methodology. For completeness, we will briefly present both here.[4]

[3] In a discrete setting, for example, multidimensional types can always be reassigned to a one-dimensional parameter, but the single-crossing property is not always preserved.

[4] Most recent methodological treatments of the screening problem use the parametric-utility approach, referred to by Wilson (1993a) as the "disaggregated-type" approach. See, for example, the article by Guesnerie and Laffont (1984), and the relevant sections in Fudenberg and Tirole (1991), Myerson (1991), Mas-Colell, Whinston, and Green (1995), and Stole (1997). The demand-profile approach is thoroughly expounded in Wilson (1993a). Brown and Sibley (1986) and Wilson (1993a) discuss both approaches.

2.1. The Parametric-Utility Approach

The basic methodology we follow here was initially developed by Mirrlees (1971), and applied to nonlinear pricing by Mussa and Rosen (1978). The firm in our setting cares only about expected profit and so seeks to maximize

$$E[\pi] = \int_{\underline{\theta}}^{\overline{\theta}} [P(q(\theta)) - cq(\theta)]\, dF(\theta),$$

where $q(\theta)$ is a parametric representation of the choice of type θ consumers, and c is the constant marginal cost of producing q units for a given consumer.

Suppose that our monopolist offers a nonlinear, lower-semi-continuous pricing schedule $P(q)$ defined over the compact domain $\mathcal{Q}$. Then, we can define a type θ consumer's indirect utility under this scheme as

$$u(\theta) \equiv \max_{q \in \mathcal{Q}} \{v(q, \theta) - P(q)\}.$$

Provided that the derivatives of v are bounded, $u(\theta)$ is absolutely continuous.

Applying the revelation principle, we can reparameterize our problem and focus on maximizing expected profits over all incentive-compatible and individually rational mechanisms, $\{p(\theta), q(\theta)\}_{\theta \in \Theta}$. As is well known, a mechanism in this context is incentive-compatible if and only if, for almost all θ, we have $\dot{u}(\theta) = v_\theta(q(\theta), \theta)$ and $q(\theta)$ is nondecreasing.[5] The former condition is equivalent to the local first-order condition and arises as a natural analog of the envelope condition in Roy's identity; the latter is equivalent to the local second-order condition. When preferences satisfy the single-crossing property, the local second-order condition implies a global condition as well. Hence, our monopolist firm can maximize expected profits subject to $\dot{u}(\theta) = v_\theta(q(\theta), \theta)$ and the monotonicity of q.

Given our incentive compatibility conditions are stated in terms of u and q, it is useful to transform our monopolist's program from price-quantity space to the utility-quantity space. Because $S(q, \theta) \equiv v(q, \theta) - cq$ represents joint surplus from producing q units of output to be consumed by a type θ consumer, the firm's expected profit can be restated as

$$E[\pi] = \int_{\underline{\theta}}^{\overline{\theta}} [S(q(\theta), \theta) - u(\theta)]\, dF(\theta). \tag{2.1}$$

Hence, the monopolist maximizes (2.1) over $\{q(\theta), u(\theta)\}_{\theta \in \Theta}$ subject to $\dot{u}(\theta) = v_\theta(q(\theta), \theta)$, $q(\theta)$ nondecreasing and subject to individual rationality. Note that this program is entirely defined by the social surplus function $S(q, \theta)$ and the partial derivative of the consumer's utility function with respect to θ. For example, the setting in which utility over quantity is $v(q, \theta) = \theta q - \frac{1}{2}q^2$ and cost is cq is formally equivalent to the setting in which a monopolist sells a

[5] Throughout, we use the notation $\dot{x}(y)$ to represent the derivative of x with respect to y.

product line with various qualities, in which the consumer's value of consuming one unit of quality q is given by $\tilde{v}(q, \tilde{\theta}) = \tilde{\theta} q$ and the cost of producing such a unit is $\frac{1}{2}q^2$, where $\tilde{\theta} = \theta - c$. Both settings give rise to identical surplus functions and partial derivatives with respect to type, and hence have identical optimal price schedules. In this sense, there is little to distinguish the use of quality [as in Mussa and Rosen's (1978) seminal paper] from quantity [as in Maskin and Riley's (1984) generalization of this model]. Fortunately, in both cases, the operative instruments begin with the letter q and, as a pleasant historical accident, we can refer to this second-best allocation as the MR allocation. We will nonetheless focus our attention on the quantity variation of this model.

As a technical simplification, we use the local first-order condition for truth-telling to replace u in the firm's objective via integration by parts. The result is an objective function that is maximized over $\{\{q(\theta)\}_{\theta \in \Theta}, u(\underline{\theta})\}$ subject to q nondecreasing and the individual rationality constraint:

$$E[\pi] = \int_{\underline{\theta}}^{\overline{\theta}} \left(S(q(\theta), \theta) - \frac{1 - F(\theta)}{f(\theta)} v_\theta(q(\theta), \theta) - u(\underline{\theta}) \right) dF(\theta).$$

This objective function has been usefully termed the monopolist's "virtual surplus" function by Myerson (1991); it includes the total surplus generated by the monopolist's production less the information rents that must be left to the consumers as a function of their type.

Because $\dot{u}(\theta) = v_\theta(q, \theta) \geq 0$, individual rationality is equivalent to requiring $u(\underline{\theta}) \geq 0$. Thus, we choose $u(\underline{\theta}) = 0$ as a *corner solution* in this program, guaranteeing participation at the least possible cost. Note that, in this simple program, it is never profitable to leave excess rents to consumers. Hence, we are left with an objective function that can be maximized pointwise in $q(\theta)$ if we ignore the monotonicity condition. Providing that the virtual surplus

$$\Lambda(q, \theta) \equiv S(q, \theta) - \frac{1 - F(\theta)}{f(\theta)} v_\theta(q, \theta)$$

is quasi-concave in q and satisfies a cross-partial condition, $\Lambda_{q\theta} \geq 0$, the solution $\{q(\theta)\}_{\theta \in \Theta}$, which is defined by the pointwise first-order condition $\Lambda_q(q(\theta), \theta) = 0$, maximizes expected profit and is nondecreasing as required. This solution satisfies

$$S_q(q(\theta), \theta) = \frac{1 - F(\theta)}{f(\theta)} v_\theta(q(\theta), \theta) \geq 0.$$

Hence, we have the familiar result that $q(\theta)$ is distorted downward relative to the social optimum, everywhere but at the "top" (i.e., at $\theta = \overline{\theta}$). If $\Lambda_{q\theta}$ is not everywhere nonnegative, it is possible that an ironing procedure needs to be used to constrain $q(\theta)$ to be nondecreasing.[6] Such a procedure typically requires that we utilize more general control-theoretic techniques and depart from our

[6] See, for example, Fudenberg and Tirole (1991) for details.

simple pointwise maximization program. However, in the single-dimensional setting, a mild set of regularity conditions on v and F guarantees us the simple case.[7]

Note that because profit per customer, $\pi = S - u$, is linear in utility, we are able to use integration by parts to eliminate the utility function from the objective function, except for the requirement that $u(\underline{\theta}) \geq 0$. This allows us to maximize profits pointwise in q; i.e., we do not have to concern ourselves simultaneously with the value of $u(\theta)$. In this sense, the program is block recursive: first the optimum can be found for each $q(\theta)$ and for $u(\underline{\theta})$ in isolation; then using the resulting function $q(\theta)$ and $u(\underline{\theta})$, $u(\theta)$ can be determined via integration. The resulting utility schedule can then be combined with $q(\theta)$ to determine the total type-specific transfer, $p(\theta) = v(q(\theta), \theta) - u(\theta)$. Given $\{p(\theta), q(\theta)\}_{\theta \in \Theta}$, the price schedule can be constructed by inverting the function $q(\theta)$: $P(q) = p(\theta^{-1}(q))$.

A second inherent simplicity in the one-dimensional model is that the incentive compatibility conditions are determined by a simple differential equation and a monotonicity condition. Whether we use integration by parts or the maximum principle to solve the program, in both instances we made important use of this fact: without it, we also lose the recursive nature of the problem. In the multidimensional setting, if we are uncertain as to which constraints bind, we will generally be forced to maximize profits subject to a far larger set of global constraints.

To this end, it is useful to briefly consider the discrete setting. Suppose that θ is distributed discretely on $\underline{\theta} = \theta_1 < \theta_2 < \cdots < \theta_I = \overline{\theta}$, with respective probabilities $f_i > 0$, $i = 1, \ldots, I$ and cumulative distribution function $F_k \equiv \sum_{i=1}^{k} f_i$. A direct mechanism is a menu of I price-quantity pairs, where the ith indexed pair is given to consumers who report that they are of the ith type: $\{q_i, p_i\}_{i=1,\ldots,I}$. Given that the single-crossing property is satisfied, it is straightforward to show that, if adjacent incentive compatibility constraints are satisfied, then global incentive compatibility is satisfied. The adjacent constraints are typically referred to as the *downward local* and *upward local* incentive constraints:

$$v(q_i, \theta_i) - p_i \geq v(q_{i-1}, \theta_i) - p_{i-1}, \quad \text{for} \quad i = 2, \ldots, I, \qquad (\text{IC}_{i,i-1})$$

$$v(q_i, \theta_i) - p_i \geq v(q_{i+1}, \theta_i) - p_{i+1}, \quad \text{for} \quad i = 1, \ldots, I-1. \qquad (\text{IC}_{i,i+1})$$

Furthermore, assuming that it is always profitable to transfer rents from the consumer to the firm, one can easily demonstrate that the downward constraints are binding. In addition, providing that the resulting quantity allocation, $\{q_i\}_{i=1,\ldots,I}$, is monotonic, one can show that the upward constraints must be slack and consequently incentive compatibility is global. This set of results is typically used to solve the relaxed program with only the downward constraints. In this sense, the sequence of binding downward-local incentive constraints (and the difference

[7] The commonly made assumptions that preferences are quadratic, and θ has a log-concave distribution are sufficient for $\Lambda_{q\theta} \geq 0$.

equation that they imply) are analogous to the ordinary differential equation $\dot{u}(\theta) = v_\theta(q(\theta), \theta)$ in the continuous setting. Not surprisingly, the solution to the relaxed program (ignoring monotonicity constraints) satisfies an analogous condition:

$$S_q(q_i, \theta_i) = \left(\frac{1 - F_i}{f_i}\right)\{v_q(q_i, \theta_{i+1}) - v_q(q_i, \theta_i)\},$$
$$i = 1, \ldots, I - 1.$$

In the discrete setting case, it is perhaps easier to see the importance of focusing on the local constraints, and in particular on the downward-local constraints. Without such a simplification, we would have to introduce a Lagrange multiplier for every type-report pair, (i, j), resulting in $I(I - 1)$ total constraints rather than simply $I - 1$. Not only does the single-crossing property in tandem with a one-dimensional type space allow us to reduce the set of potential constraints by a factor of I, it also renders these local constraints in an tractable fashion: a simple first-order difference equation. The absence of such a convenient ordering is the source of much difficulty in the multiple-dimension setting.

2.2. The Demand-Profile Approach

An alternative approach to modeling optimal screening contracts in parametric-utility space is to work with a less primitive and perhaps more economically salient structure – demand curves ordered by type and then aggregated into a demand profile.[8] Because demand curves entirely capture consumers' preferences, there is no informational loss from restricting our attention to demand profiles. Given that they are generally easier to estimate empirically, this primitive has arguably more practical appeal. For our purposes, however, the demand profile approach is useful also in that this method more clearly illustrates the simplicity and recursiveness of the single-type framework, and also underscores the aspects of the multiple-type framework that will lead to genuine economic difficulties rather than merely technical concerns.

Consider first the continuous parameterization of demand curves that we will index by θ: an individual of type θ has a demand curve given by

$$p = v_q(q, \theta).$$

The single-crossing property is equivalent to the requirement that these demand curves do not intersect. In the parametric-utility approach where $v(q, \theta) = \theta q - \frac{1}{2}q^2$, this generates a simple family of parallel demand curves: $p = \theta - q$. The primitive object on which we will work, however, is the aggregate demand profile generated by calculating the measure of consumers who consume q or

[8] An interested reader is urged to consult Wilson (1993a) for a wealth of examples and insights into this approach. Wilson (1993a) builds on the work of Brown and Sibley (1986), who provide an earlier treatment of this approach.

more units of output with a price schedule, $P(q)$. Formally, we characterize this "cumulative" aggregate demand functional as

$$M[P(\cdot), q] = \text{Prob}[\theta \in \Theta \mid \arg\max_{x}\{v(x,\theta) - P(x)\} \geq q].$$

If the consumer's program is quasi-concave [which is equivalent to the requirement that the marginal price schedule, $p(q) \equiv P'(q)$, intersects the consumer's demand curve once from below], then consumer θ will demand q or more units if and only if $v_q(q,\theta)$ is not less than the marginal price, $p(q)$, which implies that the cumulative aggregative demand functional has a very simple form:

$$M[P(\cdot), q] = \text{ Prob } [v_q(q,\theta) \geq p(q)] \equiv N(p(q), q).$$

In this case, the problem is fully decomposable: The seller's program is to choose optimal marginal prices, $p(q)$, to maximize

$$N(p, q)[p - c]$$

pointwise for each q. Assuming that the monopolist's local first-order necessary condition is also sufficient, we can characterize the solution by

$$N(p(q), q) + \frac{\partial N(p(q), q)}{\partial p}[p(q) - c] = 0,$$

or in a more familiar inverse-elasticity formula

$$\frac{p(q) - c}{p(q)} = \frac{1}{\eta(p(q), q)}, \quad \text{where} \quad \eta(p, q) \equiv \frac{-p}{N}\frac{\partial N}{\partial p}.$$

Providing that the resulting marginal price schedule, $p(q)$, cuts each parameterized demand curve only once from below, this solution to the relaxed program will satisfy the agent's global incentive compatibility constraints. The resulting nonlinear price schedule in this case is $P(q) = P(0) + \int_0^q p(s)\,ds$, where the fixed fee is chosen optimally to induce participation for all consumers who generate nonnegative virtual surplus.

When the monopolist's program is not quasi-concave over p for all q, the solution is still given by the maximization over p of $N(p, q)(p - c)$, but the resulting marginal price schedule $p(q)$ may fail to be continuous in q, which gives rise to kinks in the price function P. This situation corresponds to the cases where $\Lambda_{\theta q} < 0$ and $q(\theta)$ is not strictly monotonic (bunching arises). Notice that in this case, also the demand profile approach is less difficult than the parametric-utility approach, which must resort to an ironing procedure.

The demand profile approach does not work well when the resulting price schedule cuts some demand curve twice. In this case the expression of the aggregated demand function M cannot be simplified, because it depends on the whole function P. As an illustration, consider the following numerical example. Suppose that there are three types of consumers with demand curves for the quantities given in the first three numeric columns (we normalize

Table 5.1. *The demand-profile approach: a numerical example*

Unit	θ_1	θ_2	θ_3	$p(q)$	$N(p(q), q)$	$R(q)$
1st	7	9	11	7	3	21
2nd	5	7	9	5	3	15
3rd	3	5	7	5	2	10
4th	1	3	5	3	2	6
5th	0	1	3	3	1	3
6th	0	0	1	1	1	1
Total						56

marginal cost to zero; see Table 5.1): The fourth numeric column represents the pointwise optimal price $p(q)$, obtained by maximizing revenue $pN(p, q)$ for the qth unit. The fifth column is the number of consumers purchasing that quantity (we have normalized the population to an average of one consumer of each type), and the final column represents the revenue attributed to the particular quantity level. Total revenue using nonlinear pricing is equal to 56, whereas a uniform-pricing monopolist would choose a price of 5 per unit, sell 9 units, and make a total revenue of 45. The simplicity of this method for finding the optimal price schedule is worth noting.

The local demand-profile representation sometimes falls short, however. If the zeros in the θ_1 type's demand curve were replaced by $1 - 2\varepsilon$, and the zero in the θ_2 type's demand curve was replaced by $1 - \varepsilon$, the maximum revenue for the 6th unit would be obtained by selling to all types, whereas it would still be optimal to sell the 5th unit only to θ_3 types. Thus, we would generate gaps in consumption choices for types θ_1 and θ_2 when we maximized $p(q)$ pointwise by q. Specifically, types θ_1 and θ_2 would each be directed to choose only units 1–4 and unit 6 (but to skip unit 5), which is not feasible. This candidate solution represents the failure of the local representation; specifically, the marginal demand profile $N(p, q)$ does not capture the consumer's true preferences, which are instead characterized by the full demand profile, $M[P(\cdot), q]$.

3. THE GENERAL MULTIDIMENSIONAL SCREENING PROGRAM

3.1. A General Discrete Formulation

We begin with the discrete setting, because it is perhaps most easiest to follow, relying on simple techniques of summation and optimization for the characterization of an optimum, unlike its continuous-type counterpart that makes use of more complex techniques in vector calculus and differential forms. Nonetheless, both approaches are closely related, and the conditions in the discrete setting have smooth analogs in the continuous setting. More importantly for the purposes of this survey, the rough equivalence between the two settings allows us to understand what is difficult about "multiple dimensions." To be precise, the problems arise not because of multiple dimensionality itself, but because of a commonly associated lack of exogenous type-ordering in multiple-dimensional environments. This source of the problem is clearest in the discrete setting, where it makes no sense to speak about dimensionality without simultaneously imposing structure on preferences.[9]

Let us consider now a more general version of the discrete model, where there are I distinct consumer types, and the monopolist produces n different goods: $q \in \mathbb{R}^n$. Hence, we can speak of there being n available instruments (i.e., varieties of goods exchanged for money). We make no assumptions on preferences, except for linearity in money. For the sake of consistency with the rest of the paper, we still parameterize gross utilities in the form $v(q, \theta_i)$ (where θ_i is the consumer type $i = 1, \ldots, I$), but we make no assumption on v. By convention, $q = 0$ represents "no consumption," and we normalize utility such that $v(0, \theta) = 0$ for all θ. We denote the allocation for consumer θ_i by the vector $q_i = q(\theta_i)$ and the associated utility by the scalar $u_i = u(\theta_i)$. We will use $\mathbf{q}$ to denote the $n \times I$ matrix $(q_1, \ldots, q_I)$, and $\mathbf{u}$ to denote the I-length row vector $(u_1, \ldots, u_I)$.

Using the parametric-utility approach, we represent the firm's expected profit as

$$E[\pi] = \sum_{i=1}^{I} f_i\{S(q_i, \theta_i) - u_i\},$$

to be maximized under the discrete incentive compatibility constraints, $\mathrm{IC}_{i,j}$, as defined previously in the one-dimensional case, and individual rationality constraints:

$$\forall i \quad u_i \equiv v(q_i, \theta_i) - P(q_i) \geq 0. \tag{IR$_i$}$$

The individual rationality constraints can be considered as a particular case of incentive compatibility constraints by defining a "dummy"-type θ_0, such that $v(q, \theta_0) \equiv 0$, which implies that it will always be optimal to choose $q_0 = 0$ and $P(0) = 0$.[10] The firm's problem is thus to maximize its expected profit under implementability (i.e., IC and IR) constraints:

$$\forall i, j \in \{0, \ldots, I\}, \quad u_i \geq v(q_j, \theta_i) - P(q_j),$$

or, equivalently,

$$\forall i, j \in \{0, \ldots, I\}, \quad u_i - u_j \geq v(q_j, \theta_i) - v(q_j, \theta_j). \tag{IC$_{ij}$}$$

[9] For example, whether we index preferences using two dimensions, $v(q, \theta_{1i}, \theta_{2j})$ where $i, j = 1, \ldots, I/2$, or a single dimension, $v(q, \theta_k)$ with $k = 1, \ldots, I$, is immaterial by itself. Fundamentally, any difficulty in extending single-dimensional models to multidimensional models must arise from a lack of ordering among the types rather than any primitive notions of dimensionality.

[10] This is just a convention. It is compatible with fixed fees, because P can be discontinuous in 0.

Following Spence (1980), it is natural to decompose the firm's problem in two subproblems:

1. Minimize expected utility for fixed $\mathbf{q} = (q_1, \ldots, q_I)$.
2. Choose $\mathbf{q}$ to maximize expected surplus minus expected utility.

It is remarkable that the first subproblem has a general solution that can be found by a relatively simple algorithm. Let us denote by $\mathcal{U}(\mathbf{q})$ the set of utility vectors that implement $\mathbf{q}$. That is,

$$\mathcal{U}(\mathbf{q}) = \{(u_1, \ldots, u_I) \text{such that } \mathrm{IC}_{ij} \text{ is satisfied for all } i, j = 0, \ldots, I \text{ and } u_0 = 0\}.$$

In what follows, it will be useful to consider arbitrary paths in the set $\Theta = \{\theta_1, \ldots, \theta_I\}$. We will denote such a path from type θ_i to θ_j by the function γ. We denote the "length" of γ by ℓ; i.e., ℓ is the number of segments used to connect $\theta_i = \gamma(0)$ to $\theta_j = \gamma(\ell)$. Hence, γ is a mapping, $\gamma : \{0, 1, \ldots, \ell\} \to \Theta$. Finally, we say that a path of length ℓ is "closed" if $\gamma(0) = \gamma(\ell)$. With this notation for discrete paths, the following characterization of $\mathcal{U}(\mathbf{q})$ can be stated. A proof is found in Rochet (1987).

Lemma 3.1. *$\mathcal{U}(\mathbf{q})$ is nonempty if and only if for every closed ℓ-length path γ*

$$\sum_{k=0}^{\ell-1} v(q_{\gamma(k)}, \theta_{\gamma(k+1)}) - v(q_{\gamma(k)}, \theta_{\gamma(k)}) \leq 0. \tag{3.1}$$

To provide an intuition for condition (3.1), define the *incremental utility* between type i and type j as the difference between the utility of type i and the utility of type j when consuming the bundle assigned to type j. Condition (3.1) means that, for all closed paths γ in the set of types, the sum of incremental utilities along γ is nonpositive. Consider, for example, a closed path of length k. Incentive compatibility requires

$$u_{\gamma(k+1)} - u_{\gamma(k)} \geq v(q_{\gamma(k)}, \theta_{\gamma(k+1)}) - v(q_{\gamma(k)}, \theta_{\gamma(k)}).$$

By summing over these inequalities, we see that condition (3.1) is implied by incentive compatibility for any closed path. Lemma 3.1 says that the converse is true: condition (3.1) implies incentive compatibility, as well.[11] The proof is constructive: Lemma 3.2 gives an algorithm for constructing the minimal element of $\mathcal{U}(\mathbf{q})$.

[11] The reader versed in vector calculus will recognize this as a discrete variation of the requirement that $v_\theta(q(\theta), \theta)$ is a conservative field, where $\mathcal{C}$ represents an arbitrary closed path in Θ:

$$\oint_{\mathcal{C}} v_\theta(q(\theta), \theta)\, d\theta = 0.$$

Lemma 3.2. *When (3.1) is satisfied, $\mathcal{U}(\mathbf{q})$ has a unique minimal element, $\underline{\mathbf{u}}$, characterized for $i = 0, \dots, I$ by*

$$\underline{u}_i \equiv \sup_{\gamma} \sum_{k=0}^{\ell-1} v(q_{\gamma(k)}, \theta_{\gamma(k+1)}) - v(q_{\gamma(k)}, \theta_{\gamma(k)}), \tag{3.2}$$

where the sup *is taken over all* **open** *paths from* 0 *to* i*, and* $\underline{u}_0 \equiv 0$.

Condition (3.2) means that agent i is guaranteed a utility level, $\underline{u}_i$, equal to the sum of the incremental utilities along any path connecting θ_0 to θ_i. We will refer to this ith element of the minimum of $\mathcal{U}(\mathbf{q})$ as the *informational rent* of agent i. Note that this rent does not depend on the frequencies $\{f_1, \dots, f_I\}$ of the distribution of types, but only on the support, $\Theta = \{\theta_1, \dots, \theta_I\}$, of this distribution. Formula (3.2) shows that the informational rent of each agent can be computed by a recursive algorithm. Intuitively, it is as if each type i chooses the path from θ_0 to θ_i that maximizes the sum of incremental utilities. Denote by u_i^ℓ the maximum of formula (3.2) over all paths of length *less than or equal to* ℓ from 0 to i. u_i^ℓ can be computed recursively by the Bellman-type formula:

$$u_i^{\ell+1} = \max_j \left\{ u_j^\ell + v(q_j, \theta_i) - v(q_j, \theta_j) \right\}.$$

Condition (3.1) implies that this algorithm has no cycles. The set of types being finite, u_i^ℓ converges to the rent of agent i in a finite number of steps as ℓ is increased to I.

For any allocation $\mathbf{q}$, the dynamic programming principle implies that if j belongs to the optimal path γ from 0 to i, the truncation of γ to the path between 0 and j defines the optimal path from 0 to j.

This allows us to define a partial ordering $\prec$ on types:

$$j \prec i \iff j \text{ belongs to one of the optimal paths from 0 to } i.$$

For generic[12] allocations, there is a unique optimal path γ [with $\gamma(0) = 0$ and $\gamma(\ell) = i$] from 0 to i, and the rent of i is easily computed:

$$u_i(\mathbf{q}, \prec) = \sum_{k=1}^{\ell-1} [v(q_{\gamma(k)}, \theta_{\gamma(k+1)}) - v(q_{\gamma(k)}, \theta_{\gamma(k)})].$$

Graphically, the collection of optimal paths comprises a "tree" (i.e., a connected graph without cycles such that, from the "root" vertex 0, there is a unique path to any other point in the graph); we use Γ to represent such a tree. We can therefore represent the binding incentive constraints by such a tree emanating from the type, θ_0. One can also define for all i, j, such that $i \prec j$, the "immediate successor" $s(i, j)$ of i in the direction of j by the formula

$$s(i, j) = \min\{k \mid i \prec k \prec j, k \neq i\}.$$

[12] However, the optimal allocation $\mathbf{q}$ may be such that there are several optimal paths. We give an example of such a case in Section 3.3.

Then, it is easy to see that the agent's expected rent can be written as[13]

$$\mathrm{ER}(\mathbf{q}, \prec) = \sum_{i=1}^{I} \sum_{j \succ i} f_j [v(q_i, \theta_{s(i,j)}) - v(q_i, \theta_i)].$$

In the classic one-dimensional case when the single-crossing holds, condition (3.1) reduces to the well-known monotonicity condition $q_1 \leq q_2 \leq \cdots \leq q_I$ and $\prec$ always consists of the complete ordering: $\theta_1 < \theta_2 < \cdots < \theta_I$; the associated tree is a single connected branch. In this case

$$\mathrm{ER}(\mathbf{q}, \prec) = \sum_{i=1}^{I} (1 - F_i)\,[v(q_i, \theta_{i+1}) - v(q_i, \theta_i)]\,,$$

and as previously shown in Section 2, subproblem 2 is easily solved by maximizing the virtual surplus

$$\Lambda(q_i, \theta_i) = S(q_i, \theta_i) - \frac{1 - F_i}{f_i}[v(q_i, \theta_{i+1}) - v(q_i, \theta_i)].$$

In the general case, the binding IC constraints (corresponding to the agent's optimal paths defining the tree Γ) depend on the allocation $\mathbf{q}$, which means that the virtual surplus does not have in general a simple expression. As will be illustrated later, the virtual surplus approach works only when one can anticipate a priori the optimal paths $\gamma \in \Gamma$: i.e., which IC constraints will be binding.

To summarize, from this discussion of the general discrete formulation, two conclusions emerge that are inherent in all multidimensional problems. First, and most significantly, multiple-dimension models are difficult precisely when they give rise to an endogenous ordering over the types of Θ (i.e., the set of binding IC constraints is endogenous to the choice of $\mathbf{q}$). Second, and closely related, the incentive compatibility conditions are frequently binding not only among local types, and hence the discrete analog of the first-order approach is not generally valid and a form of an integrability condition, (3.1), must necessarily be satisfied. We will see a similar structure in the continuous-type setting.

3.2. The Continuous Case

In the continuous case, the implementability condition (3.1) translates into two necessary conditions. The first is an integrability condition that requires, for every closed path $\gamma : [0, 1] \to \Theta$, that

$$\int_0^1 v_\theta(q(\gamma(s)), \gamma(s))\, d\gamma(s) = 0. \tag{3.3'}$$

[13] When i is a maximal element, the set $\{j \mid j \succ i\}$ is empty and ER does not depend on q_i.

This is equivalent to saying that $v_\theta(q(\theta), \theta)$ is the gradient[14] of some function $u(\theta)$. The second condition is a set of inequalities:

$$\forall\theta \quad D^2u(\theta) \geq v_{\theta\theta}(q(\theta), \theta), \tag{3.3''}$$

where D^2u is the Hessian matrix of any function u such that $\nabla u = v_\theta$ and the inequality is taken in the sense of matrices (i.e., $D^2u - v_{\theta\theta}$ is positive semidefinite).

The trouble is that these conditions are not sufficient, except when $v_{\theta\theta} \equiv 0$ (the linear parameterization) in which case (3.3′) and (3.3″) are necessary and sufficient for implementability by Fenchel's duality theorem[15] (Rochet, 1987).

The continuous equivalent of Lemma 3.2 is somewhat trivial. This is because the integrability condition (3.3′) implies that, for **any** path γ connecting $\gamma(0) = \theta_0$ to $\gamma(1) = \theta$, we have

$$u(\theta) = \int_0^1 v_\theta(q(\gamma(s)), \gamma(s))\, d\gamma(s).$$

Expected surplus can be computed using the divergence theorem:[16]

$$\int_\Theta u(\theta) f(\theta)\, d\theta = \int_\Theta \lambda(\theta) \cdot v_\theta(q(\theta), \theta) f(\theta)\, d\theta - \int_{\partial\Theta} \lambda(\theta) \cdot n(\theta) f(\theta) u(\theta)\, d\sigma(\theta),$$

where λ is any solution of the partial-differential equation:

$$\operatorname{div}(\lambda(\theta) f(\theta)) + f(\theta) = 0, \tag{3.4}$$

where $n(\theta)$ is the outward normal to the boundary $\partial\Theta$ of Θ, and the notation $\int_{\partial\Theta} W(\theta)\, d\sigma(\theta)$ represents the integral of some function W along the boundary $\partial\Theta$.

[14] As noticed by several authors, this is also equivalent to a set of partial differential equations reminiscent of Slutsky equations:

$$\forall n, m \quad \frac{\partial}{\partial\theta_n}\left(\frac{\partial v}{\partial\theta_m}(q(\theta), \theta)\right) = \frac{\partial}{\partial\theta_m}\left(\frac{\partial v}{\partial\theta_n}(q(\theta), \theta)\right).$$

[15] McAfee and McMillan (1988) define a Generalized Single-Crossing condition that slightly generalizes the linear case: it amounts to assuming that, for any nonlinear price, the set of types who choose the same allocation is a linear subspace. They use it to generalize the results of Laffont, Maskin, and Rochet (1987). They also find a necessary and sufficient condition for implementability.

[16] The divergence theorem is the multidimensional analog of the integration-by-parts formula. It asserts that, under regularity conditions,

$$-\int_\Theta u(\theta)\operatorname{div}[\lambda(\theta) f(\theta)]\, d\theta = \int_\Theta \lambda(\theta) \cdot \nabla u(\theta) f(\theta)\, d\theta - \int_{\partial\Theta} \lambda(\theta) \cdot n(\theta) u(\theta) f(\theta)\, d\sigma(\theta).$$

Now, the expected profit of the firm can be written as

$$E[\pi] = \int_{\Theta} \{S(q(\theta), \theta) - \lambda(\theta) v_\theta(q(\theta), \theta)\} f(\theta)\, d\theta + \int_{\partial\Theta} \lambda(\theta) \cdot n(\theta) u(\theta) f(\theta)\, d\sigma(\theta),$$

which has to be maximized under the implementability conditions (3.3′) and (3.3″). When these constraints are not binding, this problem can, in principle, be solved by point-wise maximization of virtual surplus:

$$\Lambda(q, \theta) = S(q, \theta) - \lambda(\theta) v_\theta(q, \theta).$$

The trouble is that, like in the discrete case, λ is not known explicitly. It is defined as the unique solution of partial differential equation (3.4) that satisfies the boundary condition

$$u(\theta)[\lambda(\theta) \cdot n(\theta)] = 0 \quad \text{for all} \quad \theta \text{ on } \partial\Theta.$$

It can be proved that the general solution to equation (3.4) can be computed by integrating the density, f, along arbitrary paths γ:

$$\lambda(\theta) = \int_{\gamma^{-1}(\theta)}^{1} f(\gamma(s))\, d\gamma(s).$$

Therefore, the optimal u is characterized by two objects:

- A partition of the boundary of Θ into two regions: the "lower boundary" $\partial_0\Theta$, where the participation constraint is binding [$u(\theta) = 0$] and the "upper boundary" $\partial_1\Theta$, where $\lambda(\theta) \cdot n(\theta) = 0$, which means that there is no distortion along the normal to the boundary;
- A family of paths connecting the lower boundary [where $u(\theta) = 0$] to the upper boundary (where there is no distortion). This is the continuous equivalent of the pattern found in the discrete case: a partial ordering of types along paths connecting the region where the participation constraint binds to the region where there is no distortion.

As in the discrete setting, again two ideas emerge that are distinct to the multidimensional case: (i) the set of paths connecting the lower and upper boundaries of Θ are endogenous to the choice of allocation $\{q(\theta)\}_{\theta\in\Theta}$, and (ii) an integrability condition must necessarily be satisfied.

3.3. Tractable Discrete Models

To illustrate the different patterns that can arise in multidimensional screening models and how our conclusions affect our results, we consider here a very simple example of nonlinear pricing problems, inspired by Sibley and Srinagesh

(1997) and Armstrong and Rochet (1999).[17] In those examples, a monopolist firm produces two goods $j = 1, 2$ at a constant marginal cost (normalized to zero). There are two types of consumers, characterized by independent linear inverse demands

$$p_{ij}(q_{ij}) = \theta_{ij} - q_{ij}, \quad j = 1, 2, \quad i = 1, 2.$$

Thus types are bidimensional $\theta_i = (\theta_{i1}, \theta_{i2}), i = 1, 2$. Linear demands are equivalent to quadratic utilities:

$$v(\theta_i, q_i) = \sum_{j=1}^{2} \left\{ \theta_{ij} q_{ij} - \frac{1}{2} q_{ij}^2 \right\},$$

where q_i is the vector $q_i = (q_{i1}, q_{i2})$. The first-best efficient allocation is characterized by the vector $q_i^* = \theta_i$ and surplus by the scalar $S_i^* = \frac{1}{2}(\theta_{i1}^2 + \theta_{i2}^2)$.

Following lemma 3.1, the implementability condition reduces to[18]

$$(\theta_1 - \theta_2) \cdot q_2 + (\theta_2 - \theta_1) \cdot q_1 \leq 0. \tag{3.5}$$

Providing this condition is satisfied, lemma 3.2 implies that, at the optimum, the rents to the types are given by

$$u_1 = \max(0, (\theta_1 - \theta_2) \cdot q_2),$$
$$u_2 = \max(0, (\theta_2 - \theta_1) \cdot q_1).$$

The implementability condition then implies that either u_1 or u_2 equals 0 (i.e., the IR constraint binds somewhere). To fix ideas, we assume that $S_1^* < S_2^*$. By analogy with the unidimensional case, one may conjecture that the second-best allocation is then characterized by $u_1 = 0$ (binding IR "at the bottom") and $q_2 = q_2^*$ (efficiency "at the top"). This is indeed one possible regime, illustrated by Figure 5.1.

In this first case,

$$u_2 = (\theta_2 - \theta_1) \cdot q_1$$

and

$$q_1 = \theta_1 - \frac{f_2}{f_1}(\theta_2 - \theta_1).$$

This allocation can be implemented by a menu of two-part tariffs: tariff 1 has a low fixed fee $T_1 = \frac{1}{2} q_1^2$ and a unit price vector $p_1 = f_2/f_1(\theta_2 - \theta_1)$; tariff 2 has a high fixed fee $T_2 = S_2^* - u_2$ and a zero unit price vector. Note that unit prices are not necessarily above marginal costs (which have been normalized to zero), because we did not assume[19] $\theta_2 > \theta_1$. Apart from this feature the completely

[17] Dana (1992) and Armstrong (1999a) also provide related examples of tractable discrete-type models.

[18] The only relevant closed path to consider is the cycle from θ_1 to θ_2.

[19] The case $\theta_2 > \theta_1$ corresponds to what Sibley and Srinagesh (1997) have called uniformly ordered demands.

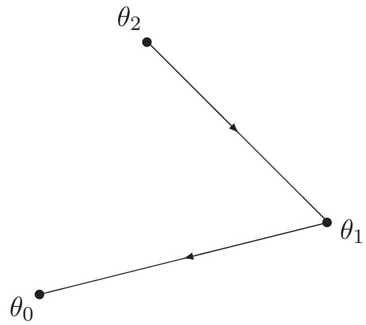

Figure 5.1. First-regime – the completely ordered case. Arrows indicate the direction of the binding incentive constraints; e.g., an arrow from θ_2 to θ_1 represents type θ_2's indifference between their own allocation and that meant for θ_1.

ordered case is analogous to the unidimensional case. It corresponds to the solution of the monopoly problem whenever

$$u_2 = (\theta_2 - \theta_1) \cdot q_1 \geq 0 = u_1 \geq (\theta_1 - \theta_2) \cdot q_2.$$

The second inequality is implied by the first, given the implementability condition in (3.5), whereas the first inequality is equivalent to

$$f_1(\theta_2 - \theta_1) \cdot \theta_1 \geq f_2(\theta_2 - \theta_1)^2,$$

or

$$\theta_1 \cdot \theta_2 \geq \frac{\theta_1^2 + f_2\theta_2^2}{1 + f_2}. \tag{3.6}$$

When this condition is not satisfied, a second possible regime corresponds to the case where there is no interaction between types; we call it the separable case (see Figure 5.2).

In this second case, there are no distortions:

$$q_1 = \theta_1 \quad \text{and} \quad q_2 = \theta_2,$$

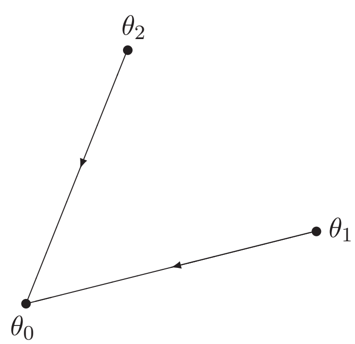

Figure 5.2. Second regime – the separable case.

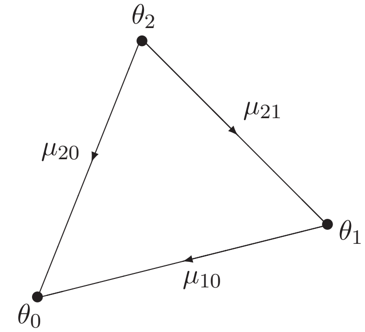

Figure 5.3. Third regime – the mixed case.

and all the surplus is captured by the seller:

$$u_1 = u_2 = 0.$$

Following (3.5), this allocation is implementable if and only if

$$(\theta_2 - \theta_1) \cdot \theta_1 \leq 0, \qquad (\theta_1 - \theta_2) \cdot \theta_2 \leq 0.$$

Given our assumption that $\theta_1^2 \leq \theta_2^2$, this is equivalent to

$$\theta_1 \cdot \theta_2 \geq \theta_1^2. \tag{3.7}$$

Finally, when neither (3.6) nor (3.7) is satisfied, there is an intermediate case that combines the features of the two regimes (see Figure 5.3).

In this third and final case, the firm is still able to capture all the surplus, but this is at the cost of a distortion on q_1, designed in such a way that type θ_2 is just indifferent between the efficient bundle $q_2 = \theta_2$ at a total tariff $T_2 = S_2^*$ and bundle q_1 at tariff $T_1 = \frac{1}{2}q_1^2$. Notice that there are two optimal paths connecting θ_2 to θ_0, corresponding to two different trees, Γ_1 and Γ_2. The weight μ_{21} put on this second path is determined by this indifference condition:

$$u_2 = 0 = (\theta_2 - \theta_1)q_1,$$

where

$$q_1 = \theta_1 - \frac{\mu_{21}}{f_1}(\theta_2 - \theta_1).$$

This gives

$$\mu_{21} = f_1 \frac{(\theta_2 - \theta_1) \cdot \theta_1}{(\theta_2 - \theta_1)^2},$$

which has to be between 0 and f_2. These conditions determine the boundary of this regime in the parameter space:

$$0 \leq (\theta_2 - \theta_1) \cdot \theta_1 \leq f_2(\theta_2 - \theta_1) \cdot \theta_2,$$

or

$$\theta_1^2 \leq \theta_1 \cdot \theta_2 \leq \frac{\theta_1^2 + f_2\theta_2^2}{1 + f_2}.$$

Notice that, in this case, we have that $u_1 = u_2 = 0$ but $q_1 \neq q_2$, which cannot arise in dimension 1.

The three cases (completely ordered, separable, and mixed) illustrate the three settings that generally arise in multidimensional models. When we place significant restrictions on preferences and heterogeneity, we can frequently obtain simpler solutions that correspond to the first two cases. We discuss these in the following section, and then consider variations on these themes in Sections 5–8. The mixed case corresponds to the more general and difficult setting we discuss in Section 9.

4. AGGREGATION AND SEPARABILITY

In this section, we explore two cases where multidimensional problems can be effectively reduced to unidimensional problems: the case of **aggregation**, where a one-dimensional sufficient statistic can be found for representing unobservable preference heterogeneity, and the case of **separability**, where the set of types can be partitioned a priori into one-dimensional subsets. In the former setting, the binding IC constraints necessarily lie in a completely ordered graph, which is known a priori and corresponds to the completely ordered case, discussed in the previous section. In the latter setting, the incentive constraints can be partitioned into an exogenously given tree that is known a priori, which corresponds to the separable case.

4.1. Aggregation

A family of multidimensional screening problems that effectively reduces to one-dimensional problems are characterized by the existence of a sufficient statistic of dimension 1 that summarizes all relevant information on unobservable heterogeneity of types and that has an *exogenously* given distribution. Let us start with a trivial example where the sufficient statistics can be found immediately. Suppose that only one good is sold ($n = 1$), but types are bidimensional ($m = 2$) and social surplus is given by

$$S(\theta, q) = (\theta_1 + \theta_2)q - \frac{1}{2}q^2.$$

It is then obvious that $\hat{\theta} \equiv \theta_1 + \theta_2$ is a one-dimensional sufficient statistic for the consumer's preferences, and the monopolist's can be solved by applying the usual techniques to the distribution of $\hat{\theta}$.

Even in this simple transformable setting, however, we can see that everything is not the same as in the canonical one-dimensional model. The primary difference is the exclusion property discovered by Armstrong (1996). Suppose, indeed, that $\theta = (\theta_1, \theta_2)$ has a bounded density on $\mathbb{R}^2_+$, or on a rectangle $[\underline{\theta}_1, \bar{\theta}_1] \times [\underline{\theta}_2, \bar{\theta}_2]$ (or any domain with a "southwest" corner). Then, it is easy to see that the density of $\hat{\theta}$, obtained by convolution of the marginals[20] of θ_1 and θ_2 tends to zero when $\hat{\theta}$ tends to the lower bound of its support. As a result, the inverse hazard rate tends to infinity, which implies the existence of an exclusion region at the bottom that would not necessarily emerge if either θ_1 or θ_2 was observable and contractible. There is an associated intuition that relates to the envelope theorem: raising prices by ε raises revenues from inframarginal buyers by a first-order amount at a loss of a second-order measure of consumers in the southwest corner of the support of types, ε^2. We will see this insight extends to more general settings; for example, Armstrong (1996) originally demonstrates this result for the separable setting discussed in the following section.[21]

It is worth noting that, whereas the aggregation technique appears trivial in our toy example, it is often more subtle and arises from a property of the market setting. For example, Biais, Martimort, and Rochet (2000) consider a market maker who sells a risky asset to a population of potential investors, characterized by two dimensions of adverse selection, $\theta = (\theta_1, \theta_2)$ (using our notation): θ_1 corresponds to the investor's fundamental information; i.e., his evaluation of the asset's liquidation value [the true liquidation value is $\theta_1 + \tilde{\varepsilon}$, where $\tilde{\varepsilon}$ is $N(0, \sigma^2)$ and independent of θ_2]; θ_2 corresponds to a sort of personal taste variable – namely, the initial position of the investor in the risky asset (his hedging needs). If he buys q units of the asset for a total price $P(q)$, the investor's final wealth is

$$\tilde{W}(q) = W_0 - P(q) + (\theta_1 + \tilde{\varepsilon})(\theta_2 + q),$$

where W_0 denotes his initial endowment of money. Assuming that the investor has constant absolute risk aversion preferences $[u(W) = -e^{-\rho W}]$, the certainty equivalent of trading q units is

$$V(q) = W_0 - P(q) + \theta_1(\theta_2 + q) - \frac{1}{2}\rho\sigma^2(\theta_2 + q)^2.$$

Thus, the net utility of trading q is given by

$$U = V(q) - V(0) = (\theta_1 - \sigma^2\theta_2)q - \frac{1}{2}\rho\sigma^2 q^2 - P(q).$$

Even though the initial screening problem is bidimensional, the simplified

[20] As noticed by Miravete (1996), the monotone hazard-rate property is preserved by convolution.

[21] Armstrong (1996) shows that the exclusion property is also true when Θ is strictly convex. However, suppose that θ is uniformly distributed on a rectangle that has been rotated 45 degrees:

$$\Theta = \{\theta \in \mathbb{R}^2, \underline{\theta} \geq \theta_1 + \theta_2 \geq \bar{\theta}, -d \leq \theta_1 - \theta_2 \leq d\}.$$

Then, it is easy to see that $\hat{\theta}$ has a uniform distribution on $[\underline{\theta}, \bar{\theta}]$, which implies that $q^*(\theta) = 2\hat{\theta} - \bar{\theta}$ and that the exclusion region vanishes when $2\underline{\theta} > \bar{\theta}$. This shows that the exclusion property discovered by Armstrong (1996) is not intrinsically related to multidimensionality, but rather to the properties of the distribution of types.

version of the problem reduces to a one-dimensional screening problem with a sufficient statistic $\hat{\theta} = \theta_1 - \rho\sigma^2\theta_2$ that aggregates the two motives for trade.

Other examples of this sort appear in Laffont and Tirole (1993) and Ivaldi and Martimort (1994). Ivaldi and Martimort (1994) study a model of competition with two dimensions of preference heterogeneity, which, given their assumptions about distributions and preferences, aggregates into a model with a one-dimensional statistic. Laffont and Tirole (1993) study regulation of multidimensional firms in a model combining adverse selection and moral hazard. By assuming that costs are observable to the regulator, they effectively transform their problem into a pure screening model, amenable to the technique presented here. In particular, when the unobservable technological parameters of the firms (their "types") are multidimensional, Laffont and Tirole find conditions, inspired by the aggregation theorems of Blackorby and Schworm (1984), under which the type vectors can be aggregated into a single number.

4.2. Separability

Wilson (1993a, 1993b) and Armstrong (1996) were the first to provide closed-form solutions to multidimensional screening models. These solutions are all of the separable type. An illustration can be given in our framework by assuming linear parametrization of surplus with respect to types $S(\theta, q) = \theta \cdot q - W(q)$, where W is convex such that $\nabla W(0) = 0$, and a density f of types that depends only on $\|\theta\|$. Consider, for example, the case where there are two goods ($m = 2$), and f is the density of a truncated normal on a quarter of a circle of center 0 and radius $R > 1$.

Wilson (1993a) and Armstrong (1996) find conditions under which the solution to the monopolist problem depends only on the distribution of types along the "rays" (i.e., the straight lines through the origin). In other words, they look for cases where the only binding IC constraints are "radial" (see the Figure 5.4).

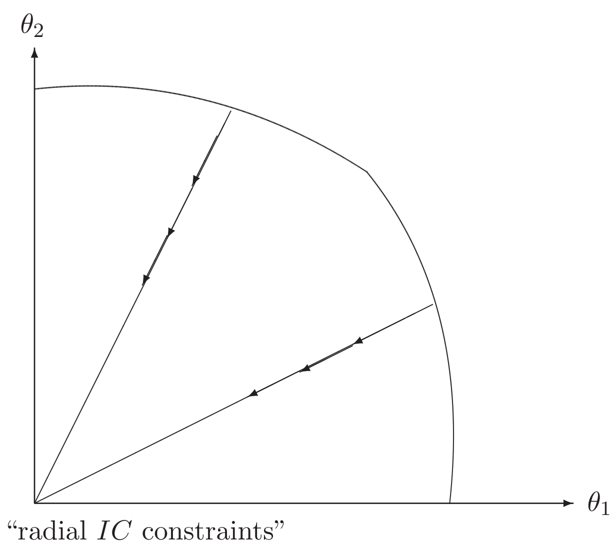

Figure 5.4. Radial incentive compatibility constraints.

If this is the case, the solution can be determined by computing the conditional distribution of types along the rays. This is done by introducing the change of variable $\theta = t\binom{\cos\alpha}{\sin\alpha}$, with $t \in [0, R]$ and $\alpha \in [0, \pi/2]$. The change of variable formula for multivariate densities gives the conditional density along the rays:

$$g(t) = t \exp -\frac{t^2}{2},$$

which does not depend on α. The virtual surplus is easily computed as

$$\Lambda(\theta, q) = S(\theta, q) - \frac{1 - G(t)}{g(t)} \frac{\theta}{\|\theta\|} \cdot q,$$

which gives, after easy computations:

$$\Lambda(\theta, q) = \left(1 - \frac{1}{\|\theta\|^2}\right) \theta q - W(q).$$

This virtual surplus is maximized for $q^*(\theta)$ defined implicitly by $\nabla W(q) = (1 - 1/\|\theta\|^2)\theta$ for $\|\theta\| \geq 1$, and $q = 0$ for $\|\theta\| < 1$. If we use the indirect surplus function $S^*(\theta) = \max_q \{\theta \cdot q - W(q)\}$, this is equivalent to: $q^*(\theta) = \nabla S^*([1 - 1/\|\theta\|^2]_+ \cdot \theta)$, where $[x]_+$ denotes $\max(0, x)$.

We now have to check whether this function q^* satisfies the necessary conditions of the monopoly problem, namely boundary conditions and implementability conditions. The boundary conditions require that the boundary of $\Theta = \mathbb{R}^2_+$ be partitioned into two regions:

- $\partial_0\Theta$, where $u(\theta) \equiv 0$ (no rent)
- $\partial_1\Theta$, where the gradient of the surplus is tangent to the boundary (no distortion at the boundary).

These two regions are represented in Figure 5.5.

Notice that the boundary condition is satisfied in $\partial_1\Theta$ only because the extreme rays are tangent to the boundary. This property would not be satisfied if the support of θ was shifted by an arbitrarily small vector. This more complex case is discussed in Section 9.2. On the other hand, Armstrong (1996) discovered a robust property of the solution, namely the existence of an exclusion region (where $u \equiv 0$): in our example, it corresponds to the region $\|\theta\| \leq 1$. This is explained by the fact that, for "regular" distributions on $\mathbb{R}^2_+$ (similar properties hold for many other domains), the conditional densities along the rays tend to zero when $\|\theta\|$ tends to zero, which implies that inverse hazard rates tend to infinity, as discussed in Section 4.1.

It remains to check that implementability conditions are satisfied. Due to the linearity of preferences with respect to θ, these implementability conditions are equivalent to saying that q^* is the gradient of a convex function (i.e., that Dq^* is a symmetric, positive definite matrix). Easy computations show that symmetry is equivalent to saying that $S^*(\theta)$ depends only on $\|\theta\|$; that is, it possesses the same type of symmetry as the density of types. If this property is satisfied, the

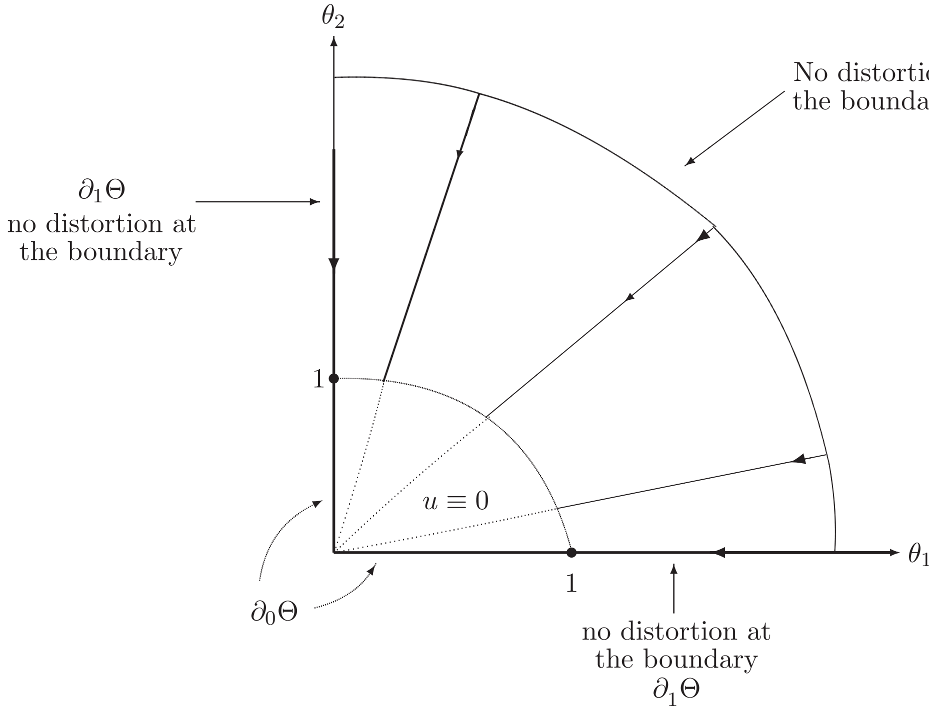

Figure 5.5. Exclusion region and boundaries.

second-order conditions for implementability (i.e., the fact that Dq^* is positive definite) will be automatically satisfied. When this is not the case, the solution is much more complex to characterize, because bunching necessarily appears. We study such an example in Section 9.2.

5. ENVIRONMENTS WITH ONE-DIMENSIONAL INSTRUMENTS

In many multidimensional screening problems, there are more dimensions of heterogeneity than instruments available to the principal ($n < m$). Here, we turn attention to the case of screening problems with one instrument ($n = 1$), but several parameters of adverse selection ($m > 1$) in which, even though a univariate sufficient statistic exists, its distribution is *endogenous*, depending on the pricing schedule chosen by the firm.

Typically, the set of instruments may be limited either by exogenous reasons [see, e.g., the justifications given by Rochet and Stole (2002) for ruling out stochastic contracts] or because the principal restricts herself to a subclass of all possible instruments. For example, Armstrong (1996) focuses on cost-based tariffs in his search of optimal nonlinear prices for a monopolist.[22] Using our notation, the monopolist problem in Armstrong (1996) can then be simplified

[22] Similarly, several authors e.g., Zheng (2000) and Che and Gale (1996a, 1996b), have studied score auctions, a particular subclass of multidimensional auctions in which the auctioneer aggregates bids using a prespecified scoring rule. As another example, Armstrong and Vickers (2000) consider price-cap regulation under the restriction of no lump-sum transfers.

by computing indirect utilities

$$V(y,\theta) = \max_q \{v(q,\theta) \mid C(q) \le y\}$$

representing the maximum utility attained by a consumer of type θ who gets a bundle of total cost less than or equal to y. The problem reduces then to find the best one-dimensional schedule $T(y)$ $(n = 1)$ for screening a multidimensional distribution of buyers $(m > 1)$.

As in the one-dimensional case, there are two approaches available for this class of problems: the parametric-utility approach and the demand-profile approach. The demand-profile approach is typically far easier to implement, provided that the consumer's preferences can be accurately summarized by a demand profile that depends only on the marginal prices.

Laffont, Maskin, and Rochet (1987) solved such a problem using the parametric-utility approach. Consider the scenario in which a monopolist sells only one good $(n = 1)$ to buyers differing by two characteristics: the intercept θ_1 and the slope $-\theta_2$ of their (individual) inverse demand curves. This corresponds to the following parameterization of preferences:

$$v(q,\theta) = \theta_1 q - \frac{1}{2}\theta_2 q^2.$$

If we want to apply the parametric-utility methodology, we are confronted with the problem that implementability of an indirect utility function $u(\cdot)$ is more complex to characterize. Indeed, let $P(q)$ be a given price schedule. The corresponding indirect utility u and allocation rule q satisfy

$$u(\theta) = \max_q \left\{\theta_1 q - \frac{1}{2}\theta_2 q^2 - P(q)\right\},$$

where the maximum is attained for $q = q(\theta)$. By the envelope principle, we have that u is again a convex function such that

$$\nabla u(\theta) = \begin{pmatrix} q(\theta) \\ -\frac{1}{2}q^2(\theta) \end{pmatrix} \quad \text{for} \quad \text{a.e. } \theta.$$

This shows that u necessarily satisfies a nonlinear partial-differential equation

$$\frac{\partial u}{\partial \theta_2} + \frac{1}{2}\left(\frac{\partial u}{\partial \theta_1}\right)^2 = 0. \tag{5.1}$$

The monopolist's problem can then be transformed as before into a calculus of variations problem in u and ∇u, but with the additional constraint (5.1) that makes the program difficult.

Interestingly, Wilson's demand-profile approach works very well in this case. Let us define the demand profile for quantity q at marginal price p as

$$N(p,q) = \text{Prob}[v_q(q,\theta) \ge p] = \text{Prob}[\theta_1 - \theta_2 q \ge p].$$

Assuming a constant marginal cost c, the optimal marginal price $p(q) = P'(q)$ can be obtained by maximizing $(p - c)N(p,q)$ with respect to p. If θ_1

and θ_2 are distributed independently according to cumulative distributions F_1 and F_2 (and densities f_1 and f_2), we obtain

$$N(p, q) = \int_0^{+\infty} \{1 - F_1(p + \theta_2 q)\} f_2(\theta_2)\, d\theta_2.$$

The optimal marginal price is defined implicitly by

$$p(q) = c - \frac{N(p(q), q)}{N_p(p(q), q)} = c + \frac{\displaystyle\int_0^{+\infty} \{1 - F_1(p(q) + \theta_2 q)\} f_2(\theta_2)\, d\theta_2}{\displaystyle\int_0^{+\infty} f_1(p(q) + \theta_2 q) f_2(\theta_2)\, d\theta_2},$$

which generalizes the classical formula obtained when θ_2 is nonstochastic:

$$p(q) = c + \frac{1 - F_1}{f_1}(p(q) + \theta_2 q).$$

For example, when θ_1 is exponentially distributed (i.e., $f_1(\theta_1) = \lambda_1 e^{-\lambda_1 \theta_1}$), the mark-up is constant and the two formulas coincide: $p(q) = c + 1/\lambda_1$.

Notice also that $\hat{\theta} = \theta_1 - \theta_2 q(\theta)$ is a univariate sufficient statistic, but unlike the case considered in Section 4.1, its distribution depends on $q(\theta)$ and thus on the price schedule chosen by the monopolist.

We now turn to a subset of these models with a single instrument, in which one dimension of type enters utilities additively.

6. ENVIRONMENTS WITH RANDOM PARTICIPATION

6.1. A General Framework

We consider a class of environments in which $n = 1$, but in which a particular additivity assumption provides sufficient structure to produce some general economic conclusions. Specifically, suppose that $n = 1$ and $m = 2$, but utility of the agent is restricted to the form

$$u = v(q, \theta_1) - \theta_2 - P,$$

where $\Theta_1 = [\underline{\theta}_1, \overline{\theta}_1]$ and $\Theta_2 = \mathbb{R}_+$.

Several interesting economic settings can be studied within this model. First, we can think of the θ_2 parameter as capturing a type-dependent participation constraint. Previous work on type-dependent participation has assumed that θ_2 is a deterministic function of θ_1 (e.g., they are perfectly correlated).[23] In this sense, the framework generalizes the previous one-dimensional literature, although many of the more interesting results rely on independent distributions of θ_1 and θ_2.

[23] See, for example, Maggi and Rodriguez-Clare (1995), Lewis and Sappington (1989a, 1989b), and Jullien (2000).

Second, one can think of θ_2 as capturing a "locational cost" in a discrete-choice model of consumer behavior.[24] This allows one to extend the nonlinear pricing model of Mussa and Rosen (1978) to a more general setting, which may be important to obtain a more realistic model of consumer behavior. As an illustration, consider the predicted consumer behavior of the standard, one-dimensional model following a uniform price increase from $P(q)$ to $P(q) + \delta$: the units sold at every quality level except the lowest should remain unchanged. This is because a shift in $P(q)$ has no effect on any of the incentive compatibility conditions, since the shift occurs on both sides of the constraints. By adding the stochastic utility effect of θ_2, predicted market shares would smoothly change for all types, although perhaps more dramatically for lower types.

Third, consider the regulatory setting first discussed in Baron and Myerson (1982). There, a regulator designs an optimal mechanism for regulating a monopoly with unknown marginal cost. Suppose that, in addition, fixed costs are also private information: i.e., $C(q) = \theta_1 q + \theta_2$. Profit for the regulated firm that receives $T(q)$ as a transfer from the regulator for producing q units is $\pi = T(q) - \theta_1 q + \theta_2$, that has a one-to-one correspondence with the previous monopoly setting.[25]

Other closely related examples that we discuss in more detail include selling to liquidity-constrained buyers, where θ_2 captures the buyer's available budget, regulation of a firm in an environment with demand and cost heterogeneity, competition between oligopolists selling differentiated products with nonlinear pricing, and competition among sellers providing goods via auctions.

The key simplifications in all of these settings are twofold. First, one dimension of information enters additively. As such, q is unavailable for direct screening on this additive attribute. Second, attention is limited to deterministic[26] price schedules, $P(q)$.

[24] See Anderson, de Palma, and Thisse (1992) for a review of this large literature, and Berry, Levinsohn, and Pakes (1995) for an econometric justification of the additive specification.

[25] Rochet (1984) first solved this problem on an example with general mechanisms that rely on randomization. Applying Rochet and Stole's (2002) results to this context is appropriate in the restricted setting in which the price schedule is deterministic. In this case, Rochet and Stole (2002) show that the presence of uncertainty over fixed costs causes the optimal regulation to reduce the extent of the production distortion.

[26] Given the relevance of deterministic contracts, this may seem a reasonable restriction, a priori. In general, however, the principal may be able to do better by introducing a second screening instrument, ϕ, which represents the probability that the agent is turned away with $q = 0$. In this case, utility becomes $\phi(v(q, \theta_1) - \theta_2 - P)$ and ϕ can be used to screen different values of θ_2. On the other hand, it is without loss of generality to rule out such random mechanisms when either (i) the value θ_2 is lost by participating in the mechanism (i.e., even if $\phi = 0$), which eliminates the possibility to screen over θ_2; alternatively, (ii) if the agent can anonymously return to the principal until $\phi = 1$ is realized, the problem is stationary and the agent will continue to return until $q > 0$, and so there is no benefit to the randomization. We leave the discussion of stochastic mechanisms unresolved and simply restrict attention to deterministic price schedules remaining agnostic about the reasons.

We take the joint density to be $f(\theta_1, \theta_2) > 0$ on $\Theta_1 \times \Theta_2$, the marginal distribution of θ_1 as $f_1(\theta_1)$, and the conditional cumulative distribution function for θ_2 as $G(\theta_2 \mid \theta_1) \equiv \int_{\underline{\theta}_2}^{\theta_2} f(\theta_1, t)\, dt$. Define the indirect utility function

$$u(\theta_1) \equiv \max_{q \in Q} \; v(q, \theta_1) - P(q).$$

This indirect utility is independent of the additive component, θ_2, because it does not affect the optimal choice of q, conditional on $q > 0$. Net utility is given by $u(\theta_1) - \theta_2$. Note that the agent's participation contains an additional random component: i.e., the agent participates iff $u(\theta_1) \geq \theta_2$. Hence, an agent with type θ_1 participates with probability $G(u(\theta_1) \mid \theta_1)$, and the expected profit of a mechanism that generates $\{q(\theta_1), u(\theta_1)\}$ for all participating agents is

$$\int_{\Theta_1} G(u(\theta_1) \mid \theta_1)\,(S(q(\theta_1), \theta_1) - u(\theta_1))\, f(\theta_1)\, d\theta_1.$$

This is maximized subject to the standard one-dimensional incentive compatibility conditions: $\dot{u}(\theta_1) = v_{\theta_1}(q(\theta_1), \theta_1)$ and $q(\theta_1)$ nondecreasing. In short, we have removed the typical corner condition that would require the utility of the lowest type – which we denote $\underline{u} \equiv u(\underline{\theta}_1)$ – to be zero, and instead introduced an endogenous determination of $\underline{u}$.

The endogeneity of $\underline{u}$ poses some difficulties that were not present in the one-dimensional setting. First and foremost, part of the block-recursive structure is now lost: There is a nonrecursive aspect to the problem as the entire function $q(\theta_1)$ and the initial condition $u(\underline{\theta}_1)$ must be *jointly* determined. Given that a purchasing consumer's preferences are ordered by a single-crossing property in (θ_1, q), the general problem of global vs. local incentive constraints is not present; incentive constraints are still recursive in their structure, although we may have to restrict q to a nondecreasing allocation. The problem is that the first-order condition determining the optimal utility for the lowest-type $\underline{u}$ depends on the optimal quantity schedule, $\{q(\theta_1)\}_{\theta_1 \in \Theta_1}$, and the first-order equation for the latter (specifically, the Euler equation) depends on the value of the former. Thus, although the resulting system of equations is not a system of partial differential equations as is common in the general multidimensional continuous type setting, but rather a second-order boundary-value problem, it is still more complicated than the standard initial-value first-order problem that arises in the canonical class of one-dimensional models.

Finding general characteristics of the solution is difficult without imposing some additional structure. A convenient restriction used in Rochet and Stole (2002) is to focus attention on independent distributions of θ_1 and θ_2, requiring that the former is distributed uniformly on Θ_1 and that the latter have a log-concave conditional cumulative distribution function.[27] Even with these distributional simplifications, the additional effect on market share still

[27] In Rochet and Stole (2002), some general results are nonetheless available in the two-type setting, providing that $G(\theta_2 \mid \theta_1)$ is log-concave in θ_2.

provides substantial difficulty. The primary cause of the difficulties is that the relaxed program (without monotonicity imposed) frequently generates nonmonotonic solutions. Hence, pooling occurs even with nonpathological distributions. Nonetheless, as a first result, one can show that if pooling occurs, it occurs only for a lower interval on Θ_1 and that otherwise efficiency occurs on the boundaries of Θ_1. This already is a substantial departure from the one-dimensional setting, and shares many similarities with the work of Rochet and Choné (1998) (see Section 9.2), especially the general presence of bunching and the efficiency on the boundaries.

Several results emerge beyond pooling or efficiency at the bottom. First, as the distribution on Θ_2 converges to an atom at $\theta_2 = 0$, the optimal allocation converges to that of the standard one-dimensional setting. Second, one can demonstrate that the optimal solution is always bounded above by the first-best allocation and below by the MR allocation. This last result has a clear economic intuition behind it. Under the standard one-dimensional setting, there is no reason not to extract the maximal amount of rent from the agents. This is the reason for distorting output downward; it allows the principal to extract greater rents from the higher types without completely shutting off the lower types from the market. When participation depends monotonically on the amount of rent left to the agent, it seems natural to leave more rents to the agent on the margin, and therefore to reduce the magnitude of the distortions. The argument is a bit more involved than this, because the presence of pooling eliminates these simple envelope-style arguments.

These results can be illustrated with a numerical example. Suppose that $\Theta_1 = [4, 5]$ and $G(\theta_2) = 1 - e^{-\theta_2/\sigma}$. Here, we use σ as a crude measure of the amount of noise in the participation constraint. As σ goes to zero, the exponential distribution converges to an atom on zero. In the example, as σ becomes small, the optimal allocation converges pointwise to the MR allocation, although pooling emerges at the bottom. For σ sufficiently large, the allocation becomes efficient on the boundaries of Θ_1 (see Figure 5.6).

Returning to our previous discussion of other applications, it should be clear that these results immediately extend to the regulatory environment of Baron and Myerson (1982), where marginal and fixed costs are represented by θ_1 and θ_2, respectively, and the regulator is restricted to offering a deterministic, nonlinear transfer schedule.

Other settings fit into this class of models in a less obvious manner. For example, consider the papers of Lewis and Sappington (1988) and Armstrong (1999a), which look at regulation of a firm in an environment of two-dimensional private information: demand is $q = x - p$, marginal cost is c, and the firm's private information is (x, c). The regulator observes only the price of the firm's output and offers a transfer that depends on price, $T(p)$. The firm's payoff is $u = (x - p)(p - c) + T(p)$; the regulator maximizes consumer surplus less transfer, $W = \frac{1}{2}(x - p)^2 - T(p)$.

Redefine the private information as $\theta_1 = x + c$ and $\theta_2 = xc$. Following similar arguments as previously described after substituting for the demand function,

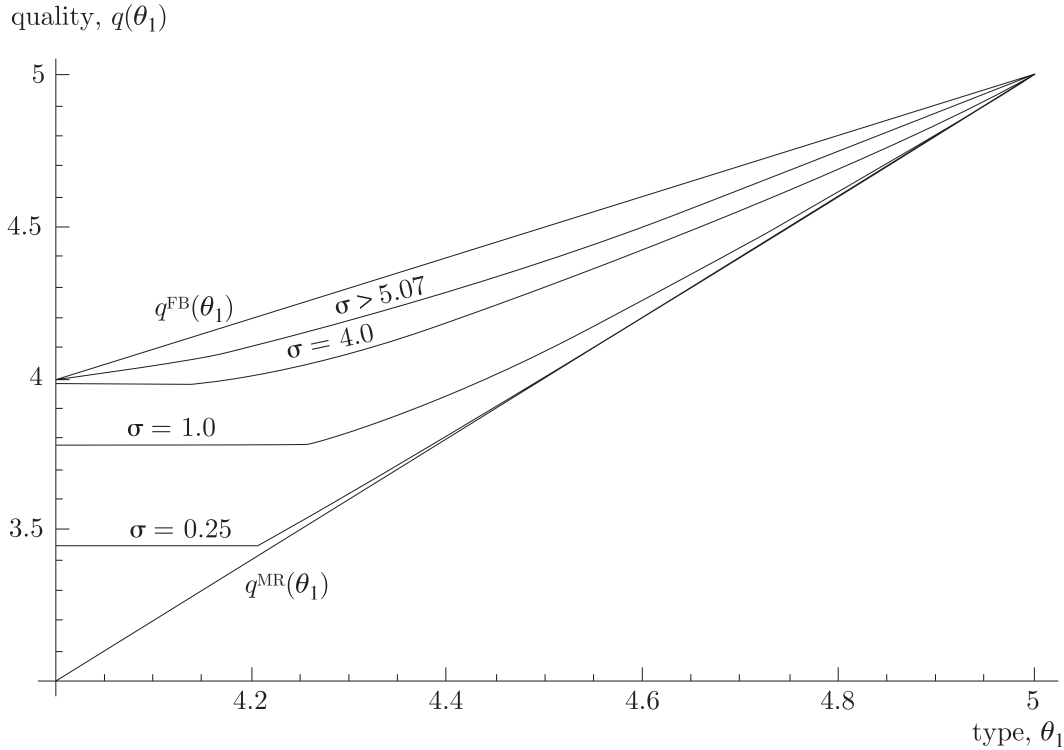

Figure 5.6. The monopoly solution with a uniform distribution of θ_1 on [4, 5] and an exponential distribution of θ_2: $G(u) = 1 - e^{-u}$.

we can define $u(\theta_1) \equiv \max_p \theta_1 p - p^2 + T(p)$ and $p(\theta_1)$ to be the corresponding maximizer. Note that the local IC constraint requires that $\dot{u}(\theta_1) = p(\theta_1) \geq 0$; second-order conditions require that u is convex in θ_1 (i.e., p is nondecreasing in θ_1). The firm will participate if and only if $u(\theta_1) \geq \theta_2$. The regulator's program can then be written as

$$\max_{\{p(\theta_1),u(\theta_1)\}} \int_{\Theta_1} G(u(\theta_1) \mid \theta_1)$$
$$\times \left[\gamma_1(\theta_1, u(\theta_1)) + p(\theta_1)\gamma_2(\theta_1, u(\theta_1)) - \frac{1}{2}p(\theta_1)^2 - u(\theta_1)\right] d\theta_1,$$

subject to $\dot{u}(\theta_1) = p(\theta_1)$ and $\ddot{u}(\theta_1) \geq 0$, where $\gamma_1(\theta_1, \theta_2) = E[\frac{1}{2}x^2 \mid \theta_1, \tilde{\theta}_2 \leq \theta_2]$ and $\gamma_2(\theta_1, \theta_2) = E[c \mid \theta_1, \tilde{\theta}_2 \leq \theta_2]$.

As another example, the work on optimal taxation is frequently concerned about leaving rents to agents that is characterized as part of the principal's objective function. Here, there is a natural connection to this class of models.

As a last example, it is worth noting the recent work of Che and Gale (2000) on two-dimensional screening when one of the dimensions is the budget constraint of the buyer. In their framework, the monopolist is selling a good to a consumer with preferences $u = \theta_1 q - P$, but with a budget constraint given by θ_2. Hence, the indirect utility function is necessarily two-dimensional:

$$u(\theta_1, \theta_2) \equiv \max_{\{q \mid P(q) \leq \theta_2\}} \theta_1 q - P(q).$$

This is a departure from the basic model presented in that θ_2 does not enter utility linearly, and monetary payments can be directly informative about the buyer's budget θ_2, because a buyer cannot pay more than he has available. Although this problem looks more complicated than the previous setting, the authors demonstrate that an optimal nonlinear pricing schedule is increasing, convex, and goes through the origin. This pins down the utility of the lowest type, $\underline{u}$; efficiency at the top determines the other boundary. Although the resulting Euler equation generates a second-order differential equation, the solution can be found analytically in many simple examples. Formally, this setting differs from the previous in that the variable θ_2 represents dissipated surplus in the case of Rochet and Stole (2002), but θ_2 represents a constraint on how much money can be transferred to the principal in Che and Gale (2000). This minor difference nonetheless translates into a significant effect on the nature of the solution: in Rochet and Stole (2002) the determination of the participation region is more difficult than in Che and Gale's (2000) setting, where the latter are able to demonstrate that the optimal tariff goes through the origin and generates full participation, albeit with distorted consumption.[28]

Finally, it is worth pointing out that the general class of problems contained in this section are closely related to models of nonlinear pricing with income effects. As Wilson (1993a) has noted in his discussion of income effects (i.e., in models in which the marginal utility of money is not constant, but varies either with wealth levels or with some related parameterization), in general the Euler conditions for optimality will consist of second-order differential equations (rather than first-order in the canonical case) and fixed fees may be part of an optimal pricing schedule. Using the demand-profile approach, suppose that the income effect is modeled by a nonlinearity in money:

$$N[P, p, q] = \text{Prob}[\theta \in \Theta \mid \text{MRS}(q, I - P(q), \theta) \geq p(q)].$$

[28] One may be tempted to solve the budget-constrained class of problems in Che and Gale (2000) by appealing to the aggregation results presented earlier. In particular, a natural candidate for a sufficient statistic when there are unit demands, $q \in [0, 1]$, is $\theta = \min\{\theta_1, \theta_2\}$. This line of reasoning is flawed because $\min\{\theta_1, \theta_2\}$ is not a sufficient statistic for the consumer's marginal rate of substitution between money and q. A simple example from Che and Gale (2000) demonstrates this most clearly. Suppose first that the consumer's valuation for the good is distributed uniformly on $\Theta_1 = [0, 1]$ and the consumer's wealth is nonstochastic and equal to $\theta_2 = 2$. The revenue-maximizing unit price is $P(1) = \frac{1}{2}$ and expected revenues are $\frac{1}{4}$. Utilizing a price-quantity schedule cannot further increase revenues. Now, suppose instead that the consumer's valuation is fixed at $\theta_1 = 2$, but wealth is a random variable distributed uniformly on $\Theta_2 = [0, 1]$. In this case, $\min\{\theta_1, \theta_2\}$ is identical as in the former setting, but now the monopolist can raise expected revenues by charging the price schedule $P(q) = 2q$ for $q \in [0, 1]$. Each consumer of type θ_2 purchases the fraction $q = \theta_2/2$, and expected revenues are $\frac{1}{2}$. Aggregation fails because the marginal rates of substitution differ across the two settings and are not functions of the same aggregate statistic. In the first, the marginal rate of substitution of q for money is θ_1 if the total purchase price is less than or equal to 2 and 0 if the total price is greater than 2. In the second setting, the marginal rate of substitution is 2 if the total price is less than or equal to θ_2, and 0 otherwise.

Here, I represents income and the demand profile depends on the marginal price, $p(q)$, and the total price level, $P(q)$, since the latter affects the marginal rate of substitution of q for money. The Euler equation is

$$\frac{\partial N}{\partial P}[p(q) - C'(q)] - \frac{d}{dq}\left\{N + \frac{\partial N}{\partial p}[p(q) - C'(q)]\right\} = 0.$$

Because the second component is totally differentiated by q, a second-order differential equation arises.

The problem loses much of its tractability because N now depends on the total price level P as well as marginal price, p. Economically, the problem is complicated because the choice of a marginal price for some q will shift the consumer's demand curve via an income effect, which will affect the optimality of other marginal prices. Hence, the program is no longer block-recursive in structure as in Rochet and Stole (2002). As one raises the marginal price of a given level of output, one also lowers the participation rate for all consumers who consume that margin or greater. It is not a coincidence that, in some models of self-selection, private information over income, and exponential utility, the nature of the optimal allocation resembles that of the allocations in the nonlinear pricing context with random participation, as in Salanié (1990) and Laffont and Rochet (1998).

7. COMPETITIVE ENVIRONMENTS

This section builds on the previous sections by applying various models to study the effects of competition on the design of screening contracts. There have been some limited attempts to model imperfect competition between firms competing with nonlinear prices within a one-dimensional framework. This, for example, is the approach taken in the papers by Spulber (1989), Ivaldi and Martimort (1994), and Stole (1995). Similarly, in most work on common agency in screening environments (e.g., Stole 1991 and Martimort 1992, 1996), the agent's private information is of one dimension. Unfortunately, as argued previously, competitive models naturally suggest at least two dimensions of heterogeneity; so, the robustness of these approaches may be called into question.

Several papers have considered competitive nonlinear pricing using a variety of methodologies. We briefly survey a few papers using the demand-profile methodology with some limited success. We then present a specific form of bidimensional heterogeneity that has been successful in applied work.

7.1. A Variety of Demand-Profile Approaches

Wilson (1993a, Chapter 12) surveys the basic economics of firms competing with nonlinear prices, outlining two general classes of models. The first category supposes that there is some product differentiation between the firms. As before, an aggregate demand profile can be constructed that measures the proportion of consumers who buy from firm i at least q units when the marginal price is p;

this demand profile obviously depends on the nonlinear price schedules offered by the other firms. The first-order conditions for optimality now include terms capturing the flux of consumer purchases on the boundaries, but also isolate a competitive externality. Wilson numerically solves two models of this sort.

A second category of models discussed by Wilson (1993a) assumes that products are homogeneous. Now, to avoid the outcome of zero-profit, marginal-cost pricing between the competing firms, one has to assume some sort of extensive form game (e.g., a Cournot game where output is brought to market and then is subsequently priced with nonlinear price schedules, etc.). Several games are considered with a variety of strategic restrictions and results in Oren, Smith, and Wilson (1982) using the demand-profile approach.

7.2. A Specific Approach: Location Models (Hotelling Type)

The third, more recent, approach has been to model competition in multidimensional environments in which simple aggregation is not available by introducing one dimension of uncertainty to handle the differentiation between firms (e.g., brand location and, more generally, "horizontal" heterogeneity) and another dimension to capture important characteristics of consumer tastes that may be similar in effect across all firms (e.g., marginal willingness to pay for quantity/quality and, more generally, "vertical" heterogeneity). Recent papers that take this approach include Armstrong and Vickers (1999), Biglaiser and Mezzetti (1999), Rochet and Stole (2002), and Schmidt-Mohr and Villas-Boas (1999), among others. We briefly survey the model and results in Rochet and Stole (1997, 2002) before remarking on the similar treatments by other authors.

As we suggest, this framework for modeling oligopoly markets is quite general; we need only posit some distribution of horizontal preferences.[29] What is fundamental is that our proposed model affords both a vertical preference parameter along the lines of Mussa and Rosen (1978), while also incorporating a measure of imperfect competition by allowing for distinct horizontal preferences.[30]

For simplicity, consider the case of two firms competing on either ends of a market with unit length and transportation cost σ. We will let $\theta_2^L \equiv \theta_2$ denote the distance from a consumer located at θ_2 to the left firm and $\theta_2^R \equiv 1 - \theta_2$ denote the distance from the same consumer to the right firm. Preferences are as before: For a consumer of type (θ_1, θ_2) consuming from firm j, an amount q_j at a price

[29] Such a framework has been usefully employed recently by Laffont, Rey, and Tirole (1998a, 1998b) and Dessein (1999) for studying competition between telecommunications networks.

[30] This modeling of competition is in the spirit of some recent empirical work on price discrimination. Leslie (1999), for example, in his study of Broadway theater ticket pricing finds it useful to incorporate heterogeneous valuations of outside alternatives to capture the presence of competing firms while maintaining a distinct form of vertical heterogeneity (in this case, income) to capture variation in preferences over quality. Because Leslie (1999) takes the quality of theater seats as fixed, he does not solve for the optimal quality-price schedule. Similarly, Ginsburgh and Weber (1996) use a Hotelling-type model to study price discrimination in the European car market.

of P_j, the consumer obtains utility of $\theta_1 q_j - \theta_2^j - P_j$. We further assume that θ_1 is distributed independently of θ_2, with $F(\theta_1)$ and $G(\theta_2)$ representing the distribution of types, respectively. Each firm simultaneously posts a publicly observable price schedule, $P_i(q_i)$, after which each consumer decides which firm (if any) to visit and which price-quality pair to select. The market share of firm j among consumers of type (θ_1, θ_2) can be computed easily:

$$M_j(u_j, u_k) = G_j\left(\min\left\{\frac{u_j}{\sigma}, \frac{1}{2} + \frac{u_j - u_k}{2\sigma}\right\}\right). \tag{7.1}$$

This comes from the fact that the marginal consumer of firm j is located at a distance that is the minimum of $u_j(\theta_1)/\sigma$ (which occurs when the total market shares are less than one – the *local monopoly* regime) and $\frac{1}{2} + (u_j - u_k)/2\sigma$ (which occurs when all the market is served – the *competitive* regime). Again, using the dual approach, we can write the total expected profit of firm i as a functional involving the consumers' rents $u_i(\cdot)$ and $u_j(\cdot)$ taken as the strategic variables of the two firms:

$$u_j(\theta_1) \equiv \max_q \; \theta_1 q - P_j(q),$$

where P_j is the price schedule chosen by firm j. We obtain

$$B_j(u_j, u_k) = \int_{\underline{\theta}_1}^{\bar{\theta}_1} \{S(t, q_j(t)) - u_j(t)\} M_j(u_j(t), u_k(t))\, dt, \tag{7.2}$$

where q_i is again related to u_i by the first-order differential equation $\dot{u}_j(\theta_1) = q_j(\theta_1)$.

We now look for a Nash equilibrium of the normal form game defined by (7.1) and (7.2), where the strategy spaces of the firms have been restricted to u_i consistent with nondecreasing quality allocations. This turns out to be a difficult task in general, because of the monotonicity conditions [remember that $q_L(\cdot)$ and $q_R(\cdot)$ have to be nondecreasing]. However, if we neglect these monotonicity conditions (which can be checked ex post), competitive nonlinear prices can be characterized by a straightforward set of Hamiltonian equations.

A numerical example is illustrative. Consider, for example, the case when θ_1 is uniformly distributed on $[4, 5]$, which is shown in Figure 5.7. For σ sufficiently large (i.e., $\sigma > 14.8$), the market shares of the two firms do not adjoin: each firm is in a (local) monopoly situation and the quality allocation is exactly the same as in our previously analyzed monopoly setting.[31] Interestingly, when the market shares are adjoining for high θ_1 (i.e., $u_L(\overline{\theta}_1) + u_R(\overline{\theta}_1) \geq \sigma$), but not all θ_1 (i.e, $u_L(\underline{\theta}_1) + u_R(\underline{\theta}_1) < \sigma$), the qualitative pattern of the solution remains identical (cf. Figure 5.7 below for $\sigma = 10$). However, when $u_L(\underline{\theta}_1) + u_R(\underline{\theta}_1) \geq \sigma$ (the fully competitive regime), it turns out that quality distortions disappear completely (cf. Figure 5.7 below, $\sigma < 16/3$). In this particular case, the equilibrium pricing schedules are cost-plus-fee schedules.

[31] It can be proved that this local monopoly solution involves full separation.

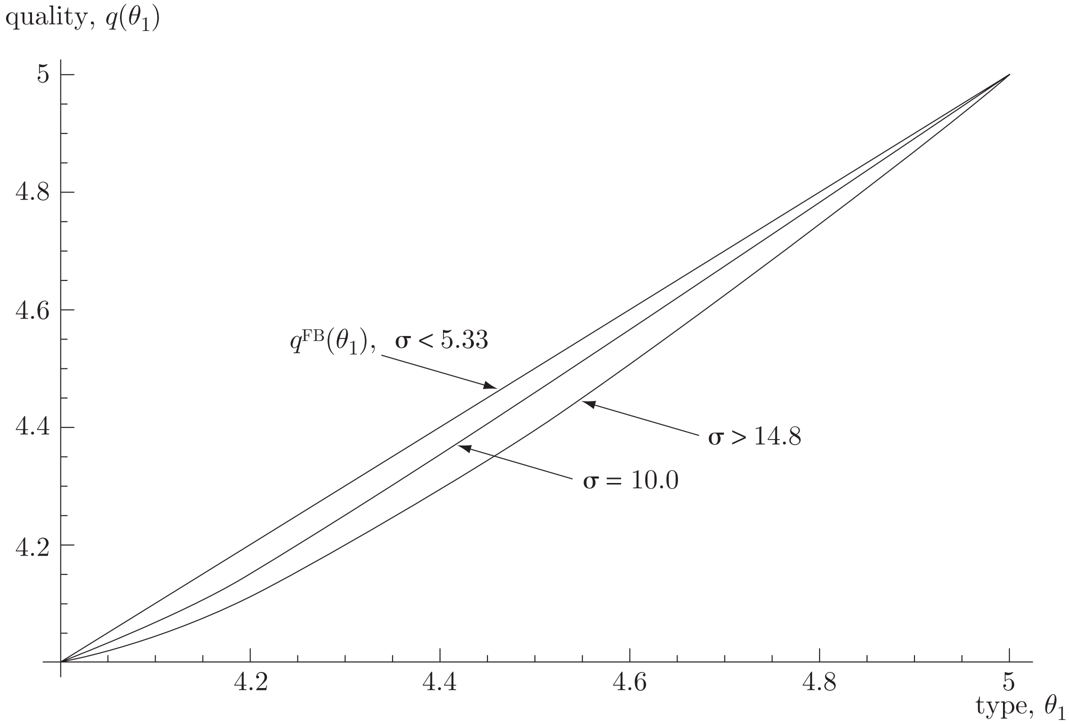

Figure 5.7. Quality choices in the oligopoly equilibrium for three regimes: fully competitive ($\sigma < 16/3$), mixed ($\sigma = 10$), and local monopoly ($\sigma > 14.8$). We assume that θ_2 is uniform on [0, 1]; θ_1 is uniform on [4, 5].

As demonstrated by Armstrong and Vickers (1999) and Rochet and Stole (2002), the result that efficiency in q emerges for a fully covered market is somewhat general, extending well beyond this simple example. Formally, if σ is sufficiently small so as to guarantee that every consumer in $\Theta_1 \times \Theta_2$ purchases, in equilibrium each firm offers a cost-plus-fee pricing schedule, $P_j(q) = C(q) + F_j$, and each customer consumes the efficient allocation $q^{fb}(t)$ from one of the two firms.

Fundamentally, this result relies on full coverage and the requirement that the inverse hazard rate is constant over θ_1 in equilibrium for each firm. More generally, we could think about an N-firm oligopoly with a joint distribution of $(\theta_2^1, \ldots, \theta_2^N)$. Formally, let $G_i(u_1, \ldots, u_N) \equiv \text{Prob}[u_i - \theta_2^i \geq \max_{j \neq i} (u_j - \theta_2^j)]$, and let the inverse hazard rate be given by

$$H_i(u_1, \ldots, u_N) = \frac{G_i(u_1, \ldots, u_N)}{\frac{\partial}{\partial u_i} G_i(u_1, \ldots, u_N)}.$$

Then, if $d/du H_i(u, \ldots, u) = 0$, for each i, cost-plus-fixed-fee pricing is an equilibrium outcome.

Biglaiser and Mezzetti (1999), in a different context, consider the case of auctions for incentive contracts of a restricted form. Because sellers have heterogeneity over their ability to provide quality, each seller's objective function takes a similar form as in Armstrong and Vickers (1999) and Rochet and Stole

(2002). Nonetheless, because of the structure of preferences and contracts, efficient cost-based pricing emerges only in the limit as preferences become homogeneous.

8. SEQUENTIAL SCREENING MODELS

A common setting in which multidimensional screening is particularly important is when information evolves over time. For example, an important paper by Baron and Besanko (1984) considers the environment in which a regulated firm learns information about its marginal cost over time, and the regulator sets prices over time as a function of the firm's sequential choices.[32] As another example, consider the problem of refund pricing studied by Courty and Li (2000). Here, a consumer purchases an airline ticket knowing that, with some probability, the value of the trip may change. The initial purchase price may depend on the refund price, particularly when marginal return to the ticket may be positively correlated with the likelihood of a change in plans or a high second-period valuation. As a third important example, Clay, Sibley, and Srinagesh (1992) and Miravete (1997) provide convincing evidence that individuals choose a variety of purchase plans, such as telephone services and electricity, which turns out ex post to be suboptimal, suggesting that the consumer is uncertain about his final needs at the time of contracting. Miravete (1997) goes on to analyze this optimal two-stage tariff problem theoretically.

In these settings, the agent learns at $t = 1$ an initial one-dimensional type parameter, θ_1, distributed according to $F_1(\theta_1)$ on Θ_1, and enters into a contractual relationship with this private information. Making an appeal to the revelation principle, without loss of generality it is assumed that the firm offers a menu of nonlinear price schedules, $\{P(q, \hat{\theta}_1)\}_{\theta_1 \in \Theta_1}$, which we index by first-period report, $\hat{\theta}_1$. Later, at $t = 2$, additional information is revealed to the agent that is economically relevant and that affects the agent's marginal utility of the contractual activity in the final period. We denote this second-period information as θ_2, which is conditionally distributed according to $F_2(\theta_2 \mid \theta_1)$ on Θ_2 with density $f(\theta_2 \mid \theta_1)$. After the realization of θ_2, the consumer chooses a particular quantity from the schedule. Assume that the consumer's final utility is given by[33]

$$u = \theta_2 q - \frac{1}{2}q^2 - P.$$

[32] Baron and Besanko (1984) also address issues of moral hazard and optimal production choice over time in the context of their model.

[33] Note that θ_2 can directly depend on θ_1 in this setting. For example, it can be the case that $\theta_2 = \theta_1 + x$, where x is independently distributed from θ_1 and is learned at $t = 2$; this is the setting studied in Miravete (1996). For conciseness, we do, however, require that the support Θ_2 is independent of θ_1.

Given the specific utility of the agent in this setting, we know that θ_2 would be a sufficient statistic for the agent's preferences, and therefore at date $t = 2$, in absence of a prior contract, there would be a simple one-dimensional problem to solve – θ_1 is payoff irrelevant conditional on θ_2.[34]

This class of models differs from previous formulations of the multidimensional problem in that the sequential nature of the information revelation restricts the agent's ability to lie. Nonetheless, the recurring theme of this survey – that multidimensional models generally pose difficulties in determining the "tree" of binding incentive constraints – reappears in the sequential context as the single-crossing property is once again endogenous. Although the papers written to date on sequential screening have typically imposed sufficient conditions to guarantee a complete ordering (i.e., the single-crossing condition), the source of the problem is still the familiar one.

To see this clearly, consider the second stage of the relationship: incentive compatibility for any given schedule chosen at $t = 1$ is guaranteed by the standard methods. Specifically, defining second-period indirect utility and optimal choice by

$$u(\hat{\theta}_1, \theta_2) \equiv \max_{q \in Q} \ \theta_2 q - \frac{1}{2} q^2 - P(q \mid \hat{\theta}_1),$$

$$q(\hat{\theta}_1, \theta_2) \equiv \arg\max_{q \in Q} \ \theta_2 q - \frac{1}{2} q^2 - P(q \mid \hat{\theta}_1),$$

second-period incentive compatibility is equivalent to $\partial u(\theta_1, \theta_2)/\partial \theta_2 = q(\theta_1, \theta_2)$ and $q(\theta_1, \theta_2)$ nondecreasing in θ_2. The approach is standard because preferences satisfy a single-crossing property in the second period. First-period incentive compatibility is more difficult to address because single crossing in (q, θ_1) is not exogenously given. Assuming second-period incentive compatibility, first-period indirect utility as a function of true type θ_1 and reported type $\hat{\theta}_1$ can be defined as

$$\tilde{u}(\hat{\theta}_1 \mid \theta_1) \equiv \int_{\underline{\theta}_2}^{\overline{\theta}_2} u(\hat{\theta}_1, \theta_2) f_2(\theta_2 \mid \theta_1)\, d\theta_2.$$

The relevant first-order local condition for truth-telling at $t = 1$ requires that

$$\frac{d}{d\theta_1} \tilde{u}(\theta_1 \mid \theta_1) = \int_{\underline{\theta}_2}^{\overline{\theta}_2} u(\hat{\theta}_1, \theta_2) \frac{\partial f_2(\theta_2 \mid \theta_1)}{\partial \theta_1}\, d\theta_2.$$

In the standard setting, the *local* second-order condition $\partial^2 \tilde{u}(\theta_1 \mid \theta_1)/\partial \theta_1 \partial \hat{\theta}_1 \geq 0$, in tandem with a monotonicity condition, yields a sufficient condition for

[34] With more general preferences, however, we could remove the presence of a one-dimensional aggregate and still find that the sequential mechanism restricts the manner in which the agent can misreport, thereby simplifying the set of binding incentive constraints.

global incentive compatibility – the global single-crossing property:

$$\partial^2 \tilde{u}(\theta_1 \mid \hat{\theta}_1)/\partial\theta_1 \partial\hat{\theta}_1 \geq 0 \ \ \forall \theta_1, \hat{\theta}_1 \in \Theta_1. \tag{scp}$$

In the present setting, this argument is not available. Instead, the focus is on maximizing the relaxed program (i.e., the program with only local first-order incentive conditions imposed), and then checking ex post that the second-order conditions are satisfied.

Substituting the agent's first-order condition into the principal's program and integrating by parts twice, we obtain the following virtual surplus for the sequential design program:

$$\Lambda(q, \theta_1, \theta_2) = \theta_2 q - \frac{1}{2}q^2 + \alpha(\theta_1, \theta_2)q - \tilde{u}(\underline{\theta}_1 \mid \underline{\theta}_1),$$

where

$$\alpha(\theta_1, \theta_2) \equiv \left(\frac{\partial F_2(\theta_2 \mid \theta_1)/\partial\theta_1}{f_2(\theta_2 \mid \theta_1)} \right) \left(\frac{1 - F_1(\theta_1)}{f_1(\theta_1)} \right).$$

Given that expected profit is $E_{\theta_1,\theta_2}[\Lambda(q, \theta_1, \theta_2)]$ and the IR constraint binds only for the lowest type $\theta_1 = \underline{\theta}_1$, profit is maximized by choosing q to maximize $\Lambda(q, \theta_1, \theta_2)$ pointwise over $\Theta_1 \times \Theta_2$. In general, the nature of the distortion will depend on the nature of the conditional distribution, $F_2(\theta_2 \mid \theta_1)$.

Baron and Besanko (1984) and Courty and Li (2000) consider the case of first-order stochastic dominance (FSD): i.e., θ_1 represents a first-order FSD shift in the distribution of θ_2. Both demonstrate that, under FSD, the IR constraint will bind for the lowest type because: (i) utility in the second period is nondecreasing in θ_2, and (ii) θ_1 shifts the distribution toward higher types. Hence, this relaxed program is cast in the appropriate form, and the principal will choose $\tilde{u}(\underline{\theta}_1, \underline{\theta}_1) = 0$. Examining the relaxed program, we see that in the FSD case, $\alpha < 0$, and the distortion in q is downward, away from the full-information solution. The intuition is by now familiar: By distorting consumption downward by a small amount, Δq, only a second-order loss in surplus arises, but a first-order gain in rent reduction is obtained as captured by $-\alpha\Delta q$. The difference in the sequential screening FSD model is that the dependence of future rents on θ_1 depends on the informativeness function, α. Baron and Besanko (1984) note the importance of commitment in this context, because the second-period allocation will be constrained efficient only if $\alpha(\theta_1, \theta_2)$ happens to equal $1 - F_2(\theta_2 \mid \theta_1)/f_2(\theta_2 \mid \theta_1)$.

When are the global incentive constraints satisfied by the q that solves the relaxed program? Baron and Besanko (1984) are silent on the sufficient conditions for global incentive compatibility, providing instead an example in which the global constraints are satisfied. Courty and Li (2000) demonstrate that if the resulting $q(\theta_1, \theta_2)$ allocation is nondecreasing in *both* arguments, then a price schedule exists that implements the allocation and satisfies global incentive compatibility.[35]

Courty and Li (2000) also consider the case in which θ_1 parameterizes the distribution of θ_2 in terms of a mean-preserving spread. Again, they demonstrate that the IR constraint will bind only for the lowest type, so the relaxed program is appropriate.[36] Global incentive compatibility in the first stage is again difficult to assess, but Courty and Li provide a useful sufficient condition to this end.[37] Interestingly, this incentive problem shares many similarities with the one-dimensional problem in which the sign of the cross-partial of expected utility with respect to q and θ_1 changes sign as one varies $(q, \theta_1) \in \mathcal{Q} \times \Theta_1$; see Araujo and Moreira (2000) for a discussion.[38] Taking the distributional assumption of Courty and Li, the solution to the relaxed program has a simple economic description. For all stage-one types, $\theta_1 < \overline{\theta}_1$, the principal introduces a *distortion in the second-period adjustment*. One can think of the final price as a markup over cost that depends on the difference between the final consumption and its expected value. The lower the initial type (i.e., the lower the noise in the second-stage marginal utility of consumption), the less valuable is the option to change consumption plans in the future. Note that variability creates higher value in expected consumption (which is why the IR constraint binds only for the lowest type, θ_1) and hence the monopolist will offer a lower price to this less consumption-valuing segment. The high types have high variability and also

[35] This follows immediately from the global sufficient condition for incentive compatibility,

$$\frac{\partial^2 \tilde{u}(\theta_1 \mid \hat{\theta}_1)}{\partial \theta_1 \partial \hat{\theta}_1} = \int_{\underline{\theta}_2}^{\overline{\theta}_2} \frac{\partial u(\hat{\theta}_1, \theta_2)}{\partial \hat{\theta}_1} \frac{\partial f_2(\theta_2 \mid \theta_1)}{\partial \theta_1} d\theta_2 \geq 0.$$

Given that $q(\theta_1, \theta_2)$ is nondecreasing in the first argument, the first term in the integrand is a nondecreasing function of θ_2. Because θ_1 represents an FSD improvement in θ_2, this integral must be nonnegative. Moreover, the fact that $q(\theta_1, \theta_2)$ is nondecreasing in its second argument guarantees second-period global incentive compatibility. Because global incentive compatibility does not require that $q(\theta_1, \theta_2)$ be nondecreasing in the first argument, when the relaxed solution is neither monotone nor globally incentive compatible, the solution to the unrelaxed program is more complex than simply using an ironing procedure on q in the relaxed program.

[36] This follows from the necessary condition for second-stage incentive compatibility: $q(\theta_1, \theta_2)$ is nondecreasing in θ_2. This (with the local first-order condition) implies that $u(\theta_1, \theta_2)$ is convex in θ_2, and hence a mean-preserving spread must increase utility; hence the IR constraint can bind only for the lowest type, θ_1.

[37] One useful simplifying assumption used by Courty and Li is that the class of distribution functions passes through the same point $\theta_2 = z$ for all θ_1. This assumption guarantees that $\alpha(\theta_1, \theta_2)$ is negative (positive) for all $\theta_2 < z$ (resp., $\theta_2 > z$). Providing that the resulting allocation $q(\theta_1, \theta_2)$ from the relaxed program is nondecreasing in each argument, global incentive compatibility is satisfied at both stages. This will be the case, for example, whenever $(\partial F_2/\partial \theta_1)/f_2$ does not vary much in θ_1.

[38] Araujo and Moreira (2000) study a one-dimensional screening model where the single-crossing condition is relaxed. As a result, nonlocal incentive constraints can be binding. They derive optimal contracts within the set of piecewise continuous contracts, and apply their techniques to bidimensional models with (perfect) negative correlation between the two dimensions of individual heterogeneity.

high option value from altering future consumption. Hence, the firm can screen these customers from the low types by charging a premium for the initial ticket, but allowing a low-cost variation in the level of final consumption. It is also the case that *any* θ_1 *type* that draws $\theta_2 = z$ in the second period will consume the efficient allocation; in our present setting, this is $q = \theta_2 = z$. The actual allocation will rotate through this point, departing from the first-best allocation $q = \theta_2$ increasingly as θ_1 decreases. Although the final allocation may have some individuals consuming above the first-best level of output, this should not be considered an *upward* distortion. Rather, the distortion is in the amount of allowed stage-two adjustment; the principal optimally distorts this adjustment downward from the efficient level.[39]

9. PRODUCT BUNDLING

In the previous discussions, we have largely focused on a variety of models that are tractable at some expense in generality. Providing, for example, that either simple aggregation or separability exists, the type space is small and discrete, or $n = 1$, we can deal with multidimensional environments with some success. We now turn to a set of models in which multidimensional screening poses the most difficult problems: $n > 1$ and $m > 1$ with nonseparable and nonaggregatable preferences. The most well-studied version of this problem is the problem of commodity bundling by a multiproduct monopolist. We will consider the papers in the literature in this context.

9.1. Some Simple Bundling Environments

We begin with the simplest linear n-product monopolist bundling environment, where $m = n$. Consumer preferences are given by

$$u = \sum_{i=1}^{n} \theta_i q_i - P,$$

where each θ_i is independently and identically distributed according to the distribution function $F(\theta_i)$ on Θ_i. (Below, we extend this model to quadratic preferences.) The cost of production is assumed to be zero, but demands are for at most one unit of each product; hence without loss of generality $q_i \in [0, 1]$.

The monopolist's space of contracts is assumed to be a price schedule $P(q_1, \ldots, q_n)$ defined on the domain $[0, 1]^n$. Given that preferences are linear in money and consumption, we can think of $q_i \in (0, 1)$ as representing either a lottery over unit consumption or partial (but deterministic) consumption. We seek to find the optimal price schedule. Nonetheless, even in this simplified setting, we are still looking for a collection of $2^n - 1$ prices.

[39] This effect is similar to that which arises in signaling models in which agents desire to signal variance to the market. See, e.g., Prendergast and Stole (1996).

Unlike the full one-dimensional ($n = m = 1$) setting in which the economics of the downward distortion is well understood, it is difficult to see the economics behind the optimal screening contract in multidimensional environments. This is in part because the multidimensional bundling environment is mathematically more complex, but also because there are at least two distinct economic effects. The first is a familiar sorting effect in which consumption is distorted downward to reduce the rents to "higher" types; the second effect arises because if demand parameters are independently distributed, a law-of-large-numbers argument shows that multigoods have a "homogenizing" effect on consumer heterogeneity. To illustrate these effects, we will present two extreme forms of this model: when $n = 2$ and when $n \to \infty$.

9.1.1. *The Case of $n = m = 2$: Similarities with the One-Dimensional Paradigm*

When $n = 2$, given the symmetry of the problem, we are looking for two marginal prices, $p(1)$ and $p(2)$; i.e., the price for one good, and the price for a second good, having already purchased the first good. The key insight is that even though the marginal values are independently distributed, the *order statistics are positively correlated*. This positive correlation makes the bundling environment akin to the classic one-dimensional paradigm. In short, provided that the first-order statistic of one consumer is greater than the first-order statistic of another, it is more likely than not that the second-order statistics are similarly ordered. Hence, it is probable that the two-good demand curves of any two consumers are nested. In this sense, a single-crossing property is present in a stochastic fashion.

To demonstrate this more precisely, denote consumer θ's first- and second-order statistics as $\theta^{(1)}$ and $\theta^{(2)}$, and refer to the corresponding first and second units of consumption as $q^{(1)}$ and $q^{(2)}$. Considering that it is physically possible to consume the second unit only after having consumed the first unit, the firm could think of this as a simple one-dimensional problem and construct the demand profile as follows:

$$N(p, q^{(i)}) = \text{Prob}[\theta^{(i)} \geq p].$$

One could then apply the one-dimensional paradigm for the demand profile to this function to obtain the optimal marginal prices. This procedure, although possibly profitable, will not obtain the maximum possible revenue. The reason why it may work in a crude sense is that a large subset of the type space will have nested demand curves (hence the one-dimensional single-crossing property will hold). Because not all types are so ordered, however, this procedure will fail to maximize revenue.

To return to the intuition for why a large subset of types are ordered as if they had one dimension, think about two consumers, where consumer A has a higher first-order statistic than consumer B. Conditional on this fact, it is also likely that consumer A will have a higher second-order statistic than B. In the case of uniformly distributed types on $\Theta_i = [0, 1]$, there is a three-fourths probability that such a nesting of demand curves will emerge between any two consumers. If the types were perfectly positively correlated, then demand curves over $\{q^{(1)}, q^{(2)}\}$ would always be nested, and we would be in the equivalent of a one-dimensional world. Because some types will have nonnested demand, a one-dimensional single-crossing property will not hold, and hence the simple demand-profile procedure will not maximize profits. Mathematically, the firm needs to account for the possibility of nonnested curves, and this alters the optimal price. A firm following the simple demand-profile procedure incorrectly perceives its profit to be

$$\pi = p(1)\text{Prob}[\theta^{(1)} \geq p^{(1)}] + p(2)\text{Prob}[\theta^{(2)} \geq p^{(2)}],$$

when in fact its profit is given by

$$\begin{aligned}\pi = {} & p(1)(\text{Prob}[\theta^{(1)} \geq p^{(1)}] + \text{Prob}[\theta^{(1)} < p^{(1)} \ \&\ \theta^{(1)} \\ & + \theta^{(2)} > p^{(1)} + p^{(2)}]) + p(2)(\text{Prob}[\theta^{(2)} \geq p^{(2)}] \\ & - \text{Prob}[\theta^{(2)} \geq p^{(2)} \ \&\ \theta^{(1)} + \theta^{(2)} < p^{(1)} + p^{(2)}]).\end{aligned}$$

There is an adjustment that must be made to the demand for each product that is not noted by the naive seller. Nonetheless, to the extent that these second terms are small, the naive one-dimensional screening approach does well in approximating the optimal solution.

This simple example makes two points. First, the well-known economic principle behind nonlinear pricing in the one-dimensional model is still present in the two-dimensional model, albeit obscured. Second, as n becomes large, the likelihood that any two consumers will have ordered demand curves decreases to zero, suggesting that the one-dimensional intuition begins to wane as n increases. Although it is difficult to make this second idea precise, we will see that a homogenizing effect along the lines of the law of large numbers removes most of the value for sorting as n increases, suggesting that the one-dimensional intuition is less appropriate for larger n.

9.1.2. *The Case of $n = m \to \infty$: The Homogenizing Effect of the Law of Large Numbers*

It has been noted in a few papers that an increase in n with independently distributed types has the effect of allowing a firm to capture most of the consumer surplus.[40] The idea is simple: selling a single aggregate bundle (again assuming marginal cost is zero) can extract most of the consumer's rents, because as n becomes large the per-unit value of this bundle converges to the sample mean. Using the argument in Armstrong (1999b), let $s_i(\theta_i) \equiv \theta_i q_i^*(\theta_i) - C(q_i^*(\theta_i))$ represent the social surplus generated by a consumer of type θ who consumes the full-information efficient allocation, $q_i^*(\theta_i)$. Suppose that the distribution of $s_i(\theta_i)$ [derived from $F(\theta_i)$] has mean μ and standard deviation σ. Then, a firm that offers cost-plus-fee pricing, $P(\mathbf{q}) = (1 - \varepsilon)\mu + C(\mathbf{q})$, where

[40] Schmalensee (1984), Armstrong (1999b), and Bakos and Brynjolfsson (1996).

$\varepsilon = 2^{\frac{1}{3}}(\sigma/\mu)^{\frac{2}{3}}$, will obtain expected profits that converge to the full-information profit level as n approaches infinity.[41] Armstrong demonstrates that this result easily extends to a setting with a particular form of positive correlation:

$$u = \theta_0 \sum_{i=1}^{n} \theta_i q_i - P.$$

Now, θ_0 is a multiplicative shock, common to all n products, but independently distributed across consumers. Armstrong shows that as n increases, the firm's profit approaches that of a monopolist with uncertainty *only* over the common component.

9.2. General Results on Product Bundling

In this section [based on Rochet and Choné (1998)], we generalize the bundling model presented to allow for multiple units demands. We come back to the general framework of nonlinear pricing by a multiproduct monopolist already studied by Wilson (1993a, 1993b) and Armstrong (1996) and presented in our Section 3.2. However, we do not assume the particular homogeneity properties of costs and types distributions that have allowed these authors to find explicit solutions using the separability property. In other words, we consider the most general multidimensional screening model in which binding IC constraints are unknown a priori. For simplicity, we assume linear-quadratic preferences:

$$u = \sum_{i=1}^{n} \left\{ \theta_i q_i - \frac{1}{2} q_i^2 \right\} - P.$$

Like before, production costs are assumed to be constant and normalized to zero, but contrary to the simple bundling model presented previously, demands for each good are not restricted to be 0 or 1. Types θ are distributed on some convex domain Θ, in accord with a continuous and positive density $f(\theta)$. Building on our previous discussions, we want to characterize the optimal pricing policy of a monopolist, using the parametric-utility approach. The problem is thus to find the function u^* that maximizes expected profit

$$E[\pi] = \int \{S(\theta, \nabla u(\theta)) - u(\theta)\} f(\theta)\, d\theta,$$

over all convex, nonnegative functions u.

When the second-order condition is not binding (i.e., when u^* is strictly convex), we already saw that u^* is characterized by two elements:

1. a partition of the boundary $\partial\Theta$ of Θ into two subsets:
   - $\partial_0\Theta$, where $u^* = 0$ (binding participation constraint), and

[41] More specifically, as shown in Armstrong (1999b), let π^* be the full-information expected profit level and let $\tilde{\pi}$ be the expected profit from the cost-plus-fee price schedule. Then, $\tilde{\pi}/\pi^*$ converges to 1 at speed $n^{-1/3}$.

- $\partial_1\Theta$, where $\partial/\partial q S(\theta, \nabla u(\theta))$ is orthogonal to the boundary of Θ (no distortion along the boundary)

2. a set of paths γ connecting $\partial_0\Theta$ to $\partial_1\Theta$, along which u^* is computed by integrating $q^*(\theta) = \nabla u^*(\theta)$.

As proved by Armstrong (1996), the nonparticipation region Θ_0 (where $u^* = 0$) typically has a nonempty interior, and u^* can be computed numerically by solving a free-boundary problem; that is, finding the curve Ψ_0 that partitions Θ into two regions: Θ_0 (where $u^* = 0$), and Θ_1, where $u^* > 0$, and, in the latter region, u^* satisfies the Euler equation:

$$\operatorname{div}\left[\frac{\partial}{\partial q} S(\theta, \nabla u(\theta) \cdot f(\theta)\right] = -f(\theta),$$

together with the boundary condition stated previously.

The problem is that, for most distributions of types [for details, see Rochet and Choné (1998)], the solution of this free-boundary problem violates the second-order conditions. The economic intuition behind this result is the presence of a strong conflict between the desire of the monopolist to limit the nonparticipation region (by pushing Ψ_0 toward the lower boundary of Θ) and "transverse" incentive compatibility constraints (that force Θ_0 and thus Ψ_0 to be convex). By trading off these two effects, the typical shape of Ψ_0 will be linear, which means that, in the region immediately above it, u^* will depend only on a linear combination of θ_1 and θ_2.

This is a robust property of multidimensional screening problems: even with log concave distributions of types, bunching cannot be ruled out, and typically occurs in the "southwest" part of Θ (i.e., for consumers with low valuations in all dimensions). From an economic viewpoint, it means that "pure bundling" (i.e., an inefficient limitation of the choice set of consumers with low valuations) is a general pattern. Rochet and Choné (1998) consider, for example, the case where θ is exponentially distributed on $[a, +\infty)^2$: with $f(\theta) = \exp(2a - \theta_1 - \theta_2)$ and $a > 1$. Because θ_1 and θ_2 are independently distributed and demands are separable, a natural candidate for the optimal price schedule is the best separable price, which can easily be computed:

$$P(q_1, q_2) = q_1 + q_2 + (a - 1)^2,$$

giving rise to demands $q_i(\theta) = \theta_i - 1$, $i = 1, 2$.

However, this *cannot* be the solution, because the nonparticipation region would be empty. In fact, the true solution has the characteristic pattern of multidimensional screening models, whereby Θ is partitioned into three regions:

- the nonparticipation region Θ_0, delimited by a first boundary Ψ_0 (of equation $\theta_1 + \theta_2 = \tau_0$) in which $u^* = 0$,
- the pure bundling region Θ_1, delimited by a second boundary Ψ_1 (of equation $\theta_1 + \theta_2 = \tau_1$) in which consumers are forced to buy a bundle with identical quantities of the two goods (thus u^* is not strictly convex, because it depends only on $\theta_1 + \theta_2$), and finally

- the fully separating region, where consumers have a complete choice and u^* can only be determined numerically.

Rochet and Choné (1998) design a specific technique, the sweeping procedure, which generalizes the ironing procedure of Mussa and Rosen (1978) for dealing with this new form of bunching, that is specific to multidimensional screening problems.

10. CONCLUDING REMARKS

In this survey, we have emphasized one general theme – that in models with multidimensional heterogeneity over preferences, the ordering of the binding incentive constraints is endogenous. Because the resulting endogenous ordering also is a source of our economic predictions, the difficulty in finding general, tractable mathematical models is particularly significant.

Notwithstanding this pessimistic appraisal, we have also emphasized in these pages that several solutions to this problem of endogenous ordering exist, all of which shed light on this issue. The simple discrete model we presented, together with a sketch of the algorithm for determining the endogenous ordering and a solution for the simple two-type case, is helpful in illustrating the economic multidimensional screening contracts – one of our primary goals in this paper. In addition, we present a variety of classes of restricted models that make the modeling tractable although still allowing sufficient theoretical degrees of freedom for interesting economics to come out of the analysis. We are particularly heartened by the recent results applied to auctions, bundling, and other rich economic settings, especially competitive environments.

ACKNOWLEDGMENTS

We are grateful to Mark Armstrong and Patrick Legros for helpful comments. The first author thanks CNRS for financial support. The second author thanks the National Science Foundation for financial support through the PFF Program. Any errors are our own.

References

Anderson, S., A. de Palma, and J.-F. Thisse (1992), *Discrete Choice Theory of Product Differentiation*. Cambridge, MA: MIT Press.

Araujo, A. and H. Moreira (2000), "Adverse Selection Problems without the Spence-Mirrlees Condition," mimeo, IMPA, Rio de Janeiro, Brazil.

Armstrong, M. (1996), "Multiproduct Nonlinear Pricing," *Econometrica*, 64(1), 51–75.

Armstrong, M. (1999a), "Optimal Regulation with Unknown Demand and Cost Functions," *Journal of Economic Theory*, 84(2), 196–215.

Armstrong, M. (1999b), "Price Discrimination by a Many-Product Firm," *Review of Economic Studies*, 66(1), 151–168.

Armstrong, M. and J.-C. Rochet (1999), "Multi-dimensional Screening: A User's Guide," *European Economic Review*, 43(4–6), 959–979.

Armstrong, M. and J. Vickers (1999), "Competitive Price Discrimination," mimeo.
Armstrong, M. and J. Vickers (2000), "Multiproduct Price Regulation under Asymmetric Information," *Journal of Industrial Economics*, 48(2), 137–159.
Bakos, Y. and E. Brynjolfsson (1996), "Bundling Information Goods: Pricing, Profits and Efficiency," Discussion Paper, MIT.
Baron, D. P. and D. Besanko (1984), "Regulation and Information in a Continuing Relationship," *Information Economics and Policy*, 1(3), 267–302.
Baron, D. and R. Myerson (1982), "Regulating a Monopolist with Unknown Costs," *Econometrica*, 50(4), 911–930.
Berry, S., J. Levinsohn, and A. Pakes (1995), "Automobile Prices in Market Equilibrium," *Econometrica* 63(4), 841–890.
Biais, B., D. Martimort, and J.-C. Rochet (2000), "Competing Mechanisms in a Common Value Environment," *Econometrica*, 68(4), 799–838.
Biglaiser, G. and C. Mezzetti (1999), "Incentive Auctions and Information Revelation," mimeo, University of North Carolina.
Blackorby, C. and W. Schworm (1984), "The Structure of Economies with Aggregate Measures of Capital: A Complete Characterization," *Review of Economic Studies*, 51, 633–650.
Brown, S. and D. Sibley (1986), *The Theory of Public Utility Pricing*. New York: Cambridge University Press.
Che, Y.-K. and I. Gale (1996a), "Expected Revenue of All-Pay Auctions and First-Price Sealed-Bid Auctions with Budget Constraints," *Economics Letters*, 50(3), 373–379.
Che, Y.-K. and I. Gale (1996b), "Financial Constraints in Auctions: Effects and Antidotes," in *Advances in Applied Microeconomics*, *Volume 6: Auctions*, (ed. by M. R. Baye), Greenwich, CT: JAI Press, 97–120.
Che, Y.-K. and I. Gale (1998), "Standard Auctions with Financially Constrained Bidders," *Review of Economic Studies*, 65(1), 1–21.
Che, Y.-K. and I. Gale (1999), "Mechanism Design with a Liquidity Constrained Buyer: The 2×2 Case," *European Economic Review*, 43(4–6), 947–957.
Che, Y.-K. and I. Gale (2000), "The Optimal Mechanism for Selling to a Budget-Constrained Buyer," *Journal of Economic Theory*, 92(2), 198–233.
Clay, K., D. Sibley, and P. Srinagesh (1992), "Ex Post vs. Ex Ante Pricing: Optional Calling Plans and Tapered Tariffs," *Journal of Regulatory Economics*, 4(2), 115–138.
Courty, P. and H. Li (2000), "Sequential Screening," *Review of Economic Studies*, 67(4), 697–718.
Dana, J. (1993), "The Organization and Scope of Agents: Regulating Multiproduct Industries," *Journal of Economic Theory*, 59(2), 288–310.
Dessein, W. (1999), "Network Competition in Nonlinear Pricing," mimeo, ECARE, Brussels.
Fudenberg, D. and J. Tirole (1991), *Game Theory*. Cambridge, MA: MIT Press, Chapter 7.
Gal, S., M. Landsberger and A. Nemirouski (1999), "Costly Bids, Rebates and a Competitive Environment," mimeo, University of Haifa.
Ginsburgh, V. and S. Weber (1996), "Product Lines and Price Discrimination in the European Car Market," mimeo.
Goldman, M. B., H. Leland, and D. Sibley (1984), "Optimal Nonuniform Prices," *Review of Economic Studies*, 51, 305–319.
Green, J. and J.-J. Laffont (1977), "Characterization of Satisfactory Mechanisms for the Revelation of Preferences for Public Goods," *Econometrica*, 45, 427–438.

Guesnerie, R. and J.-J. Laffont (1984), "A Complete Solution to a Class of Principal-Agent Problems with an Application to the Control of a Self-Managed Firm," *Journal of Public Economics*, 25, 329–369.

Hotelling, H. (1929), "Stability in Competition," *Economic Journal*, 39, 41–57.

Ivaldi, M. and D. Martimort (1994), "Competition under Nonlinear Pricing," *Annales d'Economie et de Statistique*, 34, 71–114.

Jehiel, P., B. Moldovanu, and E. Stacchetti (1999), "Multidimensional Mechanism Design for Auctions with Externalities," *Journal of Economic Theory*, 85(2), 258–293.

Jullien, B. (2000), "Participation Constraints in Adverse Selection Models," *Journal of Economic Theory*, 93(1), 1–47.

Laffont, J.-J., E. Maskin, and J.-C. Rochet (1987), "Optimal Nonlinear Pricing with Two-Dimensional Characteristics," in *Information, Incentives, and Economic Mechanisms.*, (ed. by T. Groves, R. Radner, and S. Reiter), Minneapolis MN: University of Minnesota Press.

Laffont, J.-J., P. Rey, and J. Tirole (1998a), "Network Competition: Overview and Nondiscriminatory Pricing," *Rand Journal of Economics*, 29(1), 1–37.

Laffont, J.-J., P. Rey, and J. Tirole (1998b), "Network Competition: Price Discrimination," *Rand Journal of Economics*, 29(1), 38–56.

Laffont, J.-J. and J.-C. Rochet (1998), "Regulation of a Risk-Averse Firm," *Games and Economic Behavior*, 25, 149–173.

Laffont, J.-J. and J. Tirole (1986), "Using Cost Observation to Regulate Firms," *Journal of Political Economy*, 94, 614–641.

Laffont, J.-J. and J. Tirole (1993), *A Theory of Incentives in Regulation and Procurement*. Cambridge, MA: MIT Press.

Leslie, P. (1999), "Price Discrimination in Broadway Theatre," mimeo, UCLA.

Lewis, T. and D. Sappington (1988), "Regulating a Monopolist with Unknown Demand and Cost Functions," *Rand Journal of Economics*, 19(3), 438–457.

Lewis, T. and D. Sappington (1989a), "Inflexible Rules in Incentive Problems," *American Economic Review*, 79(1), 69–84.

Lewis, T. and D. Sappington (1989b), "Countervailing Incentives in Agency Problems," *Journal of Economic Theory*, 49, 294–313.

Maggi, G. and A. Rodriguez-Clare (1995), "On Countervailing Incentives," *Journal of Economic Theory*, 66(1), 238–263.

Martimort, D. (1992), "Multi-principaux avec Anti-selection," *Annales d'Economie et de Statistique*, 28, 1–37.

Martimort, D. (1996), "Exclusive Dealing, Common Agency, and Multiprincipals Incentive Theory," *Rand Journal of Economics*, 27(1), 1–31.

Mas-Colell, A., M. Whinston and J. Green (1995), *Microeconomic Theory*. New York: Oxford University Press.

Maskin, E. and J. Riley (1984), "Monopoly with Incomplete Information," *Rand Journal of Economics*, 15, 171–196.

McAfee, R. P. and J. McMillan (1987), "Competition for Agency Contracts," *Rand Journal of Economics*, 18(2), 296–307.

McAfee, R. P. and J. McMillan (1988), "Multidimensional Incentive Compatibility and Mechanism Design," *Journal of Economic Theory*, 46(2), 335–354.

Miravete, E. (1996), "Screening Consumers Through Alternative Pricing Mechanisms," *Journal of Regulatory Economics*, 9(2), 111–132.

Miravete, E. (1997), "Estimating Demand for Local Telephone Service with Asymmetric Information and Optimal Calling Plans," Working Paper, INSEAD.

Mirrlees, J. (1971), "An Exploration in the Theory of Optimum Income Taxation," *Review of Economic Studies*, 38(114), 175–208.

Mirrlees, J. (1976), "Optimal Tax Theory: A Synthesis," *Journal of Public Economics*, 6(4), 327–358.

Mussa, M. and S. Rosen (1978), "Monopoly and Product Quality," *Journal of Economic Theory*, 18, 301–317.

Myerson, R. (1981), "Optimal Auction Design," *Mathematics of Operations Research*, 6, 58–73.

Myerson, R. (1991), *Game Theory*. Cambridge, MA: Harvard University Press.

Oren, S., S. Smith, and R. Wilson (1983), "Competitive Nonlinear Tariffs," *Journal of Economic Theory*, 29(1), 49–71.

Prendergast, C. and L. Stole (1996), "Impetuous Youngsters and Jaded Oldtimers: Acquiring a Reputation for Learning," *Journal of Political Economy*, 104(6), 1105–1134.

Rochet, J.-C. (1984), "Monopoly Regulation with Two Dimensional Uncertainty," mimeo, Université Paris 9.

Rochet, J.-C. (1987), "A Necessary and Sufficient Condition for Rationalizability in a Quasi-linear Context," *Journal of Mathematical Economics*, 16(2), 191–200.

Rochet, J.-C. and P. Choné (1998), "Ironing, Sweeping, and Multidimensional Screening," *Econometrica*, 66(4), 783–826.

Rochet, J.-C. and L. Stole (1997), "Competitive Nonlinear Pricing," mimeo.

Rochet, J.-C. and L. Stole (2000), "Nonlinear Pricing with Random Participation," *Review of Economic Studies* 69(1), 277–311.

Salanié, B. (1990), "Selection Adverse et Aversion pour le Risque," *Annales d'Economie et de Statistique*, 18, 131–149.

Schmalensee, R. (1984), "Gaussian Demand and Commodity Bundling," *Journal of Business*, 57(1), Part 2, S211–S230.

Schmidt-Mohr, U. and M. Villas-Boas (1999), "Oligopoly with Asymmetric Information: Differentiation in Credit Markets," *Rand Journal of Economics*, 30(3), 375–396.

Sibley, D. and P. Srinagesh (1997), "Multiproduct Nonlinear Pricing with Multiple Taste Characteristics," *Rand Journal of Economics*, 28(4), 684–707.

Spence, M. (1980), "Multi-Product Quantity Dependent Prices and Profitability Constraints," *Review of Economic Studies*, 47, 821–841.

Spulber, D. (1989), "Product Variety and Competitive Discounts," *Journal of Economic Theory*, 48, 510–525.

Stole, L. (1991), "Mechanism Design and Common Agency," mimeo, MIT.

Stole, L. (1995), "Nonlinear Pricing and Oligopoly," *Journal of Economics and Management Strategy*, 4(4), 529–562.

Stole, L. (1997), "Lectures on the Theory of Contracts", mimeo, University of Chicago.

Wilson, R. (1993a), *Nonlinear Pricing*. Oxford, UK: Oxford University Press.

Wilson, R. (1993b), "Design of Efficient Trading Procedures," in *The Double Auction Market: Institutions, Theories and Evidence*, Chapter 5, Santa Fe Institute Studies, Volume 14, (ed. by D. Friedman and J. Rust), Reading, MA: Addison Wesley, 125–152.

Wilson, R. (1996), "Nonlinear Pricing and Mechanism Design," in *Handbook of Computational Economics. Volume 1. Handbooks in Economics, Volume 13*, (ed. by H. Amman, D. Kendricks, and J. Rust), New York: Elsevier Science, 253–293.

Zheng, C. G. (2000), "Optimal Auction in a Multidimensional World," Discussion Paper, Northwestern University.