CARESS Working Paper 01–01

Screening Through Bundling

Eugenio J. Miravete[†⋆]

January 28, 2001

Abstract

This paper studies a class of multidimensional screening models where different type dimensions lie on the real line. The paper applies preservation results of *totally positive functions* to show that some critical properties of the distributions of asymmetric information parameters, such as increasing hazard rate, monotone likelihood ratio, and unimodality are preserved under convolution and/or composition. Under some general conditions, these preservation results also provide a natural ordering of alternative screening mechanisms. These results explain the optimality of bundling solutions in a wide range of economic models based on distributional features of the informational parameters involved. JEL: C00, D42, D82.

Keywords: Bundling, Convolution, Multidimensional Screening, Increasing Hazard Rate.

[†] Department of Economics. University of Pennsylvania, McNeil Building / 3718 Locust Walk, Philadelphia, PA 19104-6297; and CEPR, London, UK. Phone: 215-898-1505. Fax: 215-573-2057. E–mail: *miravete@ssc.upenn.edu*; http://www.ssc.upenn.edu/˜miravete

[⋆] Conversations with Jan Eeckhout, Igal Hendel, and Antonio Merlo proved very useful to widen the scope of this paper. Needless to say that I am the only responsible for any remaining errors.

1 Introduction

Multidimensional screening addresses the existence of more than a single source of heterogeneity among agents and study how the joint distribution of taste parameters condition the features of optimal contracts. Existing models have shown that dealing with multiple dimensions of agents' types leads in most cases to optimal bunching, non–monotone contracts, and even an optimal positive rent extraction for the highest consumer type. [1] In this paper I distinguish different type components but only up to their money value, *i.e.*, their combined effect on individual demand. Different type components distinguish quality dimensions of products that can be aggregated. Thus, they should be interpreted as different sources of individual demand shifts not related to price. The aggregation (addition in most cases) of type components defines a single type that is used in characterizing the bundling solution. If the process of screening the different type components is logically ordered over time, the unbundled solution is equivalent to sequential screening.

A result commonly found in the literature of multidimensional screening is that bundling is generally preferred to screening different type dimensions separately even when individual valuations are independently distributed.[2] The approach followed in this paper has the advantage that explicit, well behaved, solutions can be found for the bundled and the unbundled cases. Easy to impose regularity conditions of single–dimensional distribution of types avoid binding second order incentive compatibility constraints, and thus mechanisms are monotone and lead to fully separating equilibria. But most importantly, this setup makes possible to show that properties of the distribution of the aggregate type can be derived from those of the distributions of its components. It therefore allows to explain the principal's preference for bundling solutions as a result of well defined relations among the distributions of the type components rather than as a consequence of the complicated nonlinear optimization problems that multidimensional screening requires to solve.

The idea of aggregating different type components into a single measure is not new.[3] However, there is no systematic study in the literature on the links between the properties of distributions of type components and the aggregate type. This paper investigates how distributional features are affected by this aggregation process and offers a framework to compare alternative screening mechanisms, *i.e.*, bundling *vs.* unbundling solutions. The paper proves that under quite general conditions, common to many models, the distribution of the aggregate type θ_0, is characterized with a uniformly lower hazard rate than the distribution of any of its components θ_1 and θ_2. Intuitively, the reason is that each type dimension adds some uncertainty that increases the proportion of agents of high type. Since the distribution of aggregate types is more concentrated around high values of θ_0, the principal has to introduce higher distortions to reduce the expected

[1] See Armstrong (1996), Rochet and Choné (1998), and Wilson (1995).

[2] See Armstrong (1996, §4.6), McAfee, McMillan, and Whinston (1989), and Palfrey (1984).

[3] See Baron and Besanko (1999), Biais, Martimort, and Rochet, Miravete (1996), and most of the papers cited later in this section.

rents of inframarginal agents. Thus, bundling solutions are commonly characterized by uniformly higher markups at all purchase levels than mechanism solutions that account for informational asymmetries of each type component in isolation. Therefore, uniformly higher markups lead to higher expected profits under bundling. While the profitability ranking of screening mechanisms based on the hazard rate ordering of distributions is known since Maskin and Riley (1984), this paper shows how such hazard rate ordering may arise endogenously from the aggregation of independent type components.

The distinction between aggregate type θ_0 and type components, θ_1 and θ_2, might be useful if we want to model separately the different elements behind the optimal behavior of agents, even when all of them fall on a single real line. But at the same time, it is possible that we are more interested in the aggregate informational parameter θ_0 rather than in its components, thus reducing the dimensionality of the screening problem. Under the assumption of aggregation of type components into a single–dimensional type, we can adopt two alternative modeling approaches: we can just impose regularity conditions on preferences and distributions involving θ_0, or ensure that the combination of relevant properties of the distribution of components θ_1 and θ_2 are preserved under convolution or composition so that the solution of the model in terms of θ_0 is well behaved.

This paper focuses in this latter alternative in order to produce some useful results in a broad class of mechanism design problems. Making assumptions on the distributions of θ_1 and θ_2, instead of on the distribution of θ_0 is something that entirely depends on the nature and goal of each particular model. The results of this paper might be useful in making such a modeling decision because it shows that under fairly general conditions, the required distributional assumptions are preserved under convolution. Thus, for instance, dealing with sequential screening does not require additional assumptions on the distribution of θ_0 because they are rather implied by the distributional assumptions made on θ_1 and θ_2.

There are two key assumptions that ensure the existence of a separating equilibria in models of adverse selection. First, the single–crossing property of agents' payoff functions with respect to their control variable and the type, so that demands of different agents can be ordered for each price, and second, the increasing hazard rate (IHR) property of the distribution of types. In this paper I assume that the single–crossing property holds both for θ_0 and any of its components.[4] I will therefore focus on the necessary conditions that distributions of θ_1 and θ_2 must fulfill to ensure that the distribution of θ_0 is IHR. A similar approach will be followed to prove the preservation of the monotone likelihood ratio property (MLR).

The paper is organized as follows. Section 2 briefly reviews an apparently disperse literature to frame it according to the model of the present paper. Section 3 presents the mathematical tools needed to prove the preservation of some regularity conditions of the distributions of type components. Mathematical generality is kept to a minimum in

[4] McAffee and McMillan (1988) study the multidimensional generalization of the single–crossing property and its implications for modeling adverse selection problems.

order to prove that *total positivity* is preserved under composition and *log–concavity* is preserved under convolution. Section 4 presents the main results, *i.e.*, that IHR, MLR, and unimodality of the density functions are preserved for appropriately defined problems. This section also discusses possible extensions and applications of these preservation results to models of voting and signaling. Section 5 isolates the conditions that ensure that bundling is preferred to independent screening of type components. This section also discusses the case of correlated type components. Section 6 concludes.

2 Models of Type Aggregation

This section reviews a heterogeneous literature where an aggregate type θ_0 is defined to depend, without loss of generality, on two type components θ_1 and θ_2. The bundling solution screens agents according to the distribution of θ_0. There are more than one alternative to bundling. If type components are screened simultaneously, we have the case of multiproduct monopolists. If on the contrary these type components are screened sequentially, we encounter models of optional or sequential screening.

2.1 Multiproduct Monopolist

Consider the model of McAfee, McMillan, and Whinston (1989) where a multiproduct monopolist who must decide whether to sell his products separately or in a bundle. In both cases, and in order to reduce consumers' informational rents, the monopolist engages in nonlinear pricing. Consumers have independent taste parameters across products that define their relative intensity for them. Since these valuations are independent, their addition defines consumers' valuations of the bundle. Models that distinguish among components of a single–dimensional type can also be used in screening mechanisms involving several products. Thus, θ_1 and θ_2 are independent valuations of two components while θ_0 is the valuation of their bundle when these components are not correlated. Results presented later in Section 5 show why even when the valuations of different products are independent, bundling dominates selling these products separately as reported by McAffee, McMillan and Whinston (1989). Other contributions in this area are Adams and Yellen (1976) and Spence (1980).

Observe that the ability to aggregate different type components opens the possibility to reduce the dimensionality of the screening process. This result has been applied to models of *common agency* in Biais, Martimort, and Rochet (2000) who distinguish between θ_1, the signal of the value of an asset that is privately observed by an agent, and θ_2, the agent's endowment shock of that risky asset. Both type dimensions however aggregate into a single parameter θ_0 representing the marginal valuation of an agent for the asset to be traded, thus simplifying the design of competitive mechanisms.[5]

[5] A related paper in nonlinear pricing is the work of Sibley and Srinagesh (1997) that studies whether is is more profitable to screen the different dimensions of consumer types independently by means

A related stream of papers, such as Baron and Besanko (1999), addresses the issue of *informational alliances*, and distinguishes between unit costs of two independent suppliers, θ_1 and θ_2, and the cost profile of an informational alliance, θ_0, that can be formed to contract with the principal by consolidating the private information of agents. Other related papers in this area are Baron and Besanko (1992) and Gilbert and Riordan (1995).

Finally, besides analyzing allocation efficiency, the literature on *multi–object auctions* shows that it is generally more profitable to auction different items in a bundle than separately. In Palfrey (1983), θ_1 and θ_2 are the independent individual valuations of two objects that define the valuation of the bundle, θ_0. Armstrong (2000) shows that bundling also dominates in the case of positively correlated values in the case of optimal auctions when types can take only two values. Avery and Hendershott (2000) study whether bundling is still preferred when there are competing auctions for a subset of the products as well as when different groups of bidders have different distributions of valuations.

2.2 Sequential Screening

Consider now the following motivating example taken from Miravete (2001). A consumer wants to buy a service provided by a regulated public utility. In many occasions, the public utility offers a nonlinear tariff to maximize expected profits and reduce consumers' informational rents through the use of quantity discounts. Thus, the public utility is able to recover his fixed costs and induce efficiency gains by pricing high volume customers closer to marginal cost. Each consumer's payment is based on her particular consumption level and the shape of this single nonlinear tariff, which is critically conditioned by the assumed distribution of consumption profiles in the population. Alternatively, the public utility may offer different contract options to consumers who are now required to choose among them before their consumption is realized. Frequently, these optional contracts take the form of two–part tariffs, and are defined by a monthly fee and a particular rate per unit of consumption, although more general contracts, through a fully nonlinear tariff are also feasible, and even in some cases they might include capacity limits to consumption (*e.g.*, power load) and/or specification of the quality of the service (*e.g.*, reliability).

The difference between these two alternative pricing strategies is that the first one screens consumers based on their realized demand (bundling of type components) while the second screens consumers sequentially. In the first stage, consumers have to choose among different tariff options based on their expectation of future purchase levels. Later, once the choice have been made and individual demand is realized, each tariff option introduces different additional discounts or premia on the difference between expected and realized demand. The key feature of optional tariffs is that when consumers sign up for a particular contract option, they do not commit to any given level of consumption. At

of two–part tariffs, or alternatively by bundling all these taste parameters to design a single two–part tariff. There is however an important difference because θ_0 is no longer the sum of θ_1 and θ_2, and thus, compositions should replace convolutions to carry out the analysis.

the time of choosing tariffs, consumers are not fully aware of their own type defined as the price independent component of demand that will eventually determine the consumption level of each individual under each tariff regime. In this example, expected and actual consumption define the type of consumers. The actual type of consumers θ_0, is defined as a function of the expected consumption θ_1, and some type shock or prediction error θ_2. These are different magnitudes that can be used by the monopolist to design alternative screening mechanisms.

There are several papers that fit the described sequential screening process. For instance, in models of *expected consumption*, such as those of Ausubel (1991), Courty and Li (2000), Miravete (2000b), and Miravete (2001), individual demands are subject to independent and privately known shocks over time. The monopolist may offer a contract based on agents' actual realized demands, or alternatively a menu of optional contracts that define the payment schedule before individual demands are realized, thus taking advantage of potentially profitable effects of agent's misperception of their future consumption.

A closely related set of papers deal with the topic of *contingent pricing*. Agents differ in some idiosyncratic parameter θ_1, but their final demand is affected by the realization of some other variable θ_2, common to all consumers, that is easily observable for the principal, (*e.g.*, weather conditions). Thus, the monopolist may solve the optimal state contingent tariff that makes payment and discounts dependent on the realization of such state variable, as in Spulber (1992), or simply design a tariff that mainly target individual differences although taking into account the effect on individual demands of other variables (*e.g.*, temperature), that are not known at the time of subscribing the power capacity option, as in Panzar and Sibley (1978).

Changes in the types of agents is also a common topic in *Regulation*. The possibility of errors in the appraisal of his own cost function by the regulated firm allows regulatory agencies to consider mechanisms based either on realized or expected costs. The literature on the optimality of linear contracts –for instance in Caillaud, Guesnerie, and Rey (1992), or in Laffont and Tirole (1986)–, show that these simple contracts are robust to the existence of an additive noise θ_2 because θ_2 enters linearly in the objective function so that neither the incentive compatibility and participation constraints are changed in expectations. This is not necessarily the case in the models of *expected consumption*.[6]

Finally, uncertainty about agent's own types may also be present in *Procurement*. Awarding procurement contracts involves frequently firms bidding when they are uncertain about their future marginal costs, as in Riordan and Sappington (1987). Alternatively, the government could ask for a share of total future revenues or profits to the awarded franchisees, thus making transfers a function of actual rather than expected costs.

[6] In the latter case, the distribution of θ_2 affects the shape of the tariff based on θ_1 because it enters nonlinearly in the definition of the utility function of agents, thus making more difficult the evaluation of welfare effects. See Miravete (2001, §3.1, §6) for a detailed discussion and an empirical evaluation of these effects using data from the local telephone service industry.

3 Totally Positive and Log–Concave Functions

This section presents the minimal mathematical tools needed to prove the preservation of some key features of the distribution of asymmetric information parameters. The most important result proved in this section is that log–concavity is preserved under convolution.

ASSUMPTION 1: The random variable θ_i, $i = 1, 2$, has a continuously differentiable probability density function $f_i(\theta_i) \geq 0$ on $\Theta_i = [\underline{\theta}_i, \overline{\theta}_i] \subseteq \Re$, such that the cumulative distribution function given by:

$$F_i(\theta_i) = \int_{\underline{\theta}_i}^{\theta} f_i(z)dz, \tag{1}$$

is absolutely continuous.

Log–concavity is a smoothness property common to many distributions. It implies a certain regularity and peakedness of the density functions that makes this property very useful for the analysis of reliability. The following is a formal definition for continuously differentiable probability density functions.

DEFINITION 1: A probability distribution function $F_i(\theta_i)$ is log–concave if:

$$\frac{\partial^2 \log[f_i(\theta_i)]}{\partial \theta_i^2} = \frac{\partial}{\partial \theta_i}\left[\frac{f_i'(\theta)}{f_i(\theta)}\right] \leq 0 \quad \text{on} \quad \Theta_i. \tag{2}$$

If the distribution of an asymmetric information parameter is increasing hazard rate, then it is possible to design a screening mechanism that fully separates agents of different types, provided that the common single–crossing property of preferences holds. The IHR property is defined as follows.

DEFINITION 2: If a univariate random variable θ_i has density $f_i(\theta_i)$ and distribution function $F_i(\theta_i)$, then the ratio:

$$r_i(\theta_i) = \frac{f_i(\theta_i)}{1 - F_i(\theta_i)} \qquad \text{on} \quad \{\theta_i \in \Theta_i : F_i(\theta_i) < 1\}, \tag{3}$$

is called the hazard rate of either θ_i or $F_i(\theta_i)$. The function $\overline{F}_i(\theta_i) = 1 - F_i(\theta_i)$ is the survival function of θ_i. A univariate random variable θ_i or its cumulative distribution function $F_i(\theta_i)$ are said to be increasing hazard rate if $r_i'(\theta_i) \geq 0$ on $\{\theta_i \in \Theta_i : F_i(\theta_i) < 1\}$.

In order to infer the type of an agent from a given observable signal, models of moral hazard assume that the underlying distribution of agents' types is characterized by the monotone likelihood ratio property. This assumption is again critical to ensure the existence of separating equilibria in Principal–Agent problems characterized by the existence of moral hazard. The following is a definition of this property for continuously differentiable density functions.

Definition 3: If a univariate random variable θ_i has density function $f_i(\theta_i, \alpha)$ depending on a single indexing parameter α, then θ_i or $f_i(\theta_i, \alpha)$ are said to have the monotone likelihood ratio property if:

$$\frac{\partial^2 \ln[f_i(\theta_i, \alpha)]}{\partial\theta_i \partial\alpha} \geq 0. \tag{4}$$

In models of voting, the assumption that agents have unimodal preferences over the alternatives of the choice set becomes critical to avoid the Condorcet Paradox, the well known cyclic result in defining social preferences. The results of this paper ensure that such critical assumption is preserved if preferences are aggregated across individuals.

Definition 4: A function $f_i(\theta_i)$ is unimodal if there exists a single $\theta_i^\star \in \Theta_i$ such that $\theta_i^\star \in \arg\max_{\theta_i} f_i(\theta_i)$ in Θ_i.

In presenting the results, I will proceed from the general to the more specific version of the model. Thus, for instance, the aggregate type should be defined as any mapping of several variables into $\Re$. Again without loss of generality, I will limit the number of type components to two.

$$\theta_0 = T(\theta_1, \theta_2) : \Re^2 \rightarrow \Re. \tag{5}$$

Since θ_1 and θ_2 are random variables whose distribution is known, it is possible to characterize the distribution of the aggregate θ_0 according to equation (5). Let define the following composition operation:[7]

$$M(\theta_0, \zeta) = \int\limits_{\Theta_j} K(\theta_0, \theta_j) L(\theta_j, \zeta) dF_j(\theta_j), \tag{6}$$

where index j may take values $\{1, 2\}$. Thus, $M(\cdot)$ is a function that aggregates the dimensions θ_i and θ_j according to composition of the kernels $K(\cdot)$ and $L(\cdot)$.[8] I now have to identify the set of functions whose relevant regularity properties are preserved under composition.

Definition 5: A function $g(x, y)$ of two variables ranging over linearly ordered one–dimensional sets X and Y, respectively, is said to be *totally positive* of order n (TP_n)

[7] Kernel $K(\cdot)$ in composition equation (6), as well as $f_i(\cdot)$ in the convolution equation (13) later in the text, are normalized to integrate to one with respect to the corresponding Lebesgue measure so that they define proper density functions.

[8] The nonlinear function (5) that relates θ_0 with θ_1 and θ_2 is implicitly defined by $K(\cdot)$ and $L(\cdot)$. Thus for instance, $K(\theta_0, \theta_2) = T^{-1}(\theta_0, \theta_2)$ which expresses the probability density function of θ_1 as a function of θ_0 and θ_2 according to (5). In addition, $L(\theta_2, \zeta) = 1$ and $dF_2(\theta_2)$ defines the density of θ_0, $M(\theta_0, \zeta)$, which may also depend on the indexing parameter ζ.

if $\forall x_1 < x_2 < \ldots < x_m,\ x_i \in X \subseteq \Re$; and $\forall y_1 < y_2 < \ldots < y_m,\ y_i \in Y \subseteq \Re$; and all $1 \leq m \leq n$:

$$\begin{vmatrix} g(x_1,y_1) & g(x_1,y_2) & \cdots & g(x_1,y_m) \\ g(x_2,y_1) & g(x_2,y_2) & \cdots & g(x_2,y_m) \\ \vdots & \vdots & \ddots & \vdots \\ g(x_m,y_1) & g(x_m,y_2) & \cdots & g(x_m,y_m) \end{vmatrix} \geq 0. \tag{7}$$

The major practical significance of *totally positive functions* is that their smoothness properties (continuity, boundedness, and growth rate) are preserved under the composition operation defined in equation (6). The following Lemma states this property.[9]

LEMMA 1: *Let $K(x,y)$ and $L(x,y)$ be TP_n, and θ_1 and θ_2 be stochastically independent, then the composition:*

$$M(\theta_0,\zeta) = \int_{\Theta_2} K(\theta_0,\theta_2)L(\theta_2,\zeta)dF_2(\theta_2) = \int_{\Theta_1} K(\theta_1,\zeta)L(\theta_0,\theta_1)dF_1(\theta_1), \tag{8}$$

is also TP_n.

PROOF: Without loss of generality, let $n = 2$. By definition of TP_2, the composition $M(x,y)$ defined in (8) has to be such that $\forall x_1, x_2 \in X \subseteq \Re$ and $\forall y_1, y_2 \in Y \subseteq \Re$, such that $x_1 < x_2$ and $y_1 < y_2$, the following condition holds:

$$\begin{vmatrix} M(x_1,y_1) & M(x_1,y_2) \\ M(x_2,y_1) & M(x_2,y_2) \end{vmatrix} = \begin{vmatrix} \int K(x_1,z)L(z,y_1)dF_z(z) & \int K(x_1,z)L(z,y_2)dF_z(z) \\ \int K(x_2,z)L(z,y_1)dF_z(z) & \int K(x_2,z)L(z,y_2)dF_z(z) \end{vmatrix}$$

$$= \int_{z_1 <} \int_{z_2} \begin{vmatrix} K(x_1,z_1) & K(x_1,z_2) \\ K(x_2,z_1) & K(x_2,z_2) \end{vmatrix} \cdot \begin{vmatrix} L(z_1,y_1) & L(z_2,y_1) \\ L(z_1,y_2) & L(z_2,y_2) \end{vmatrix} dF_z(z_1)dF_z(z_2) \geq 0, \tag{9}$$

where the last inequality is the *Basic Composition Formula* that relates compositions of totally positive functions.[10] From here the proof is immediate since the first determinant in the double integral is positive as $K(x,y)$ is TP_2 and the second determinant is also positive as $L(x,y)$ is TP_2. ▪

In most economic models, when we deal with the aggregation of dimensions of agents' own types, the simple addition of the aggregate type components suffices to fully characterize consumers' types since monotone transformations of utility functions represent

[9] Observe that if $g(x,y)$ is TP_n this condition requires that all minors of order $m \leq n$ and not only the principal minors to be non–negative [Gantmacher (1958, §1.2)]

[10] The *Basic Composition Formula* is the continuous version of the Binet–Cauchy formula that expresses any minor of order k of the product of two rectangular matrices as the product of all possible minors of order k [Gantmacher (1958, §1.1)]. The proof of this intermediate result is sketched in Karlin (1968, §1.2).

the same set of preferences. However, equation (5) is justified in some other environments, such as determining the unit costs of a multiproduct firm when they are affected by production scales of two products, θ_1 and θ_2, in a specific way determined by technology. Alternatively, Courty and Li (2000) and Riordan and Sappington (1987) define θ_0 as the distribution of θ_2 conditional on the individually realized signal θ_1 which, while still keeping types single–dimensional, allows for complex interactions among type components. Transformation (6) may prove useful when dealing with the aggregation of preferences across individuals that carry some sort of weighting. However, for most of the analysis it suffices that I consider that type dimensions are related as follows:

$$\theta_0 = \theta_1 + \theta_2, \tag{10}$$

where, as before, θ_1 and θ_2 are stochastically independent, so that the distribution of the aggregate θ_0 is defined by the convolution:

$$F_0(\theta_0) = \int_{\Theta_j} F_i(\theta_0 - \theta_j) dF_j(\theta_j), \tag{11}$$

and where indices can be reversed because convolution is a commutative operation.

The structure of equations (10) – (11) captures the idea that several sources of individual heterogeneity simply translates into a single money valued magnitude that characterizes the individual reservation price of agents. Regardless of whether different type dimensions capture the effect of taste for different quality of products, the aggregation of equation (10), or more in general of equation (5), just identify the non–price driven shifts of individual demands for this product.

An important group of *totally positive functions* defines the distribution of θ_0 as the convolution of the distributions of θ_1 and θ_2 according to equations (10) – (11).[11] The set of *totally positive functions* in translation is known as *Pólya frequency functions*. The corresponding properties of convolutions of *Pólya frequency functions* are particular versions of those of composition of *totally positive functions* described above.

Definition 6: A function $g(z)$ is a *Pólya frequency function* of order n (PF_n) if $\forall x_1 < x_2 < \cdots < x_m$, $x_i \in X \subseteq \Re$; and $\forall y_1 < y_2 < \cdots < y_m$, $y_i \in Y \subseteq \Re$; and all $1 \leq m \leq n$:

$$\begin{vmatrix} g(x_1 - y_1) & g(x_1 - y_2) & \cdots & g(x_1 - y_m) \\ g(x_2 - y_1) & g(x_2 - y_2) & \cdots & g(x_2 - y_m) \\ \vdots & \vdots & \ddots & \vdots \\ g(x_m - y_1) & g(x_m - y_2) & \cdots & g(x_m - y_m) \end{vmatrix} \geq 0. \tag{12}$$

[11] Equations (5)$-$(6) reduces to the convolution case of equations (10)$-$(11) when $M(\theta_0, \zeta) = F_0(\theta_0)$, $K(\theta_0, \theta_j) = F_i(\theta_0 - \theta_j)$, and $L(\theta_j, \zeta) = 1$.

Lemma 2: *Let $f_1(\theta_1)$ and $f_2(\theta_2)$ be PF_n, and θ_1 and θ_2 be stochastically independent, then the convolution:*

$$f_0(\theta_0) = \int_{\Theta_2} f_1(\theta_0 - \theta_2) f_2(\theta_2) d\theta_2 = \int_{\Theta_1} f_1(\theta_1) f_2(\theta_0 - \theta_1) d\theta_1, \tag{13}$$

is also PF_n.

4 Results

Preservation of IHR is useful for models of adverse selection where agents' types are stochastic, thus opening the possibility to sequential screening. Similarly, preservation of MLR is useful to study agency relations in which there are several sources of moral hazard. This section proves that these preservation results are ensured by the equivalence of log–concavity to PF_2, and of MLR to TP_2 respectively.

4.1 Increasing Hazard Rate

The mathematical results of the previous section show that the smoothness properties of *Pólya frequency functions* are preserved under convolution. While reliability properties such as IHR depend on the log–concavity of the probability density functions, the preservation of such smoothness condition is easily ensured if we focus on the family of *Pólya frequency functions*. Results of this section rely on the equivalence between log–concave and a class of *Pólya frequency functions*. The following Lemma establishes this equivalence.

Lemma 3: *A continuously differentiable function $g(z)$ is PF_2 if and only if $g(z) > 0$ $\forall z \in \Re$ and $g(z)$ is log–concave on $\Re$.*

Proof: Since $g(z) > 0$ $\forall z \in \Re$, it follows from Definition 1 that a continuously differentiable function $g(z)$ is log–concave if and only if it is monotone decreasing in $\Re$. Next, without loss of generality, assume $x_1 < x_2$ and $0 = y_1 < y_2 = \Delta$. Then, from the definition of PF_2 in equation (12) and making use of common properties of determinants, the following inequality holds:

$$\lim_{\Delta\to 0} \frac{1}{\Delta} \cdot \begin{vmatrix} g(x_1) & g(x_1 - \Delta) \\ g(x_2) & g(x_2 - \Delta) \end{vmatrix} = \lim_{\Delta\to 0} \begin{vmatrix} \dfrac{g(x_1) - g(x_1 - \Delta)}{\Delta} & g(x_1 - \Delta) \\ \dfrac{g(x_2) - g(x_2 - \Delta)}{\Delta} & g(x_2 - \Delta) \end{vmatrix} = \begin{vmatrix} g'(x_1) & g(x_1) \\ g'(x_2) & g(x_2) \end{vmatrix} \geq 0, \tag{14}$$

which, given $g(z) > 0$, proves that $\forall z \in \Re$, $g'(z)/g(z)$ is monotone decreasing in $\Re$. ■

I can now prove the main result of this section. By imposing the log–concavity assumption on the probability density functions of θ_1 and θ_2, we not only identify a wide class of distributions with nice properties for economic modeling but also ensure that the distribution of θ_0 also share those properties. These results summarized in the following Proposition and Corollary.

PROPOSITION 1: *If the probability density function $f_i(\theta_i)$ is continuously differentiable and log–concave, it implies that the following properties are all equivalent:*
(a) $F_i(\theta_i)$ is log–concave,
(b) $\overline{F}_i(\theta_i) = 1 - F_i(\theta_i)$ is log–concave,
(c) $F_i(\theta_i)$ is IHR *in θ_i on $\{\theta_i \in \Theta_i : F_i(\theta_i) < 1\}$,*
(d) $l_i(\theta_i) = f_i(\theta_i)/F_i(\theta_i)$ is decreasing in θ_i on $\{\theta_i \in \Theta_i : F_i(\theta_i) > 0\}$,
(e) $f_i(\theta_i)$ is unimodal.

PROOF: See Appendix. ∎

The following Corollary shows that all the above properties are preserved under convolution, and thus, assuming that the distributions of each type component is log–concave suffices for all distributions involved to be well behaved.

COROLLARY 1: *If the probability density functions $f_i(\theta_i)$, $i = 1, 2$, are continuously differentiable and log–concave, and θ_1 and θ_2 are stochastically independent, then:*
(a) $f_0(\theta_0)$ is continuously differentiable and log–concave,
(b) $F_0(\theta_0)$ is log–concave,
(c) $\overline{F}_0(\theta_i) = 1 - F_0(\theta_i)$ is log–concave,
(d) $F_0(\theta_0)$ is IHR *in θ_0 on $\{\theta_0 \in \Theta_0 : F_0(\theta_0) < 1\}$,*
(e) $l_0(\theta_0) = f_0(\theta_0)/F_0(\theta_0)$ is decreasing in θ_0 on $\{\theta_0 \in \Theta_0 : F_0(\theta_0) > 0\}$,
(f) $f_0(\theta_0)$ is unimodal.

PROOF: By Lemma 3, $f_1(\theta_1)$ and $f_1(\theta_2)$ are both PF_2. Thus, Lemma 2 ensures that $f_0(\theta_0)$ is also PF_2. Part (a) results from applying Lemma 3 again to the convolution density function $f_0(\theta_0)$. Since the premises of Proposition 1 are now fulfilled by $f_0(\theta_0)$, parts (b)–(f) follow straightforwardly from its application. ∎

The preservation of log–concavity of distributions under convolution is the key result that ensures that a wide class of agency problems can actually be solved. Results of Proposition 1 and Corollary 1 ensure that the principal can induce separating equilibria both under the bundling and unbundling approach. The comparison of these different solutions is the goal of Section 5.

4.2 Single–Peakedness

Results of the previous section are applicable not only to screening problems. For instance, Proposition 1 and Corollary 1 also show that log–concave densities are also unimodal, and that this property is also preserved under convolution. This result is suitable to be applied to substantial issues in Political Economy since the preservation of single–peakedness of preferences is ensured.

If individual preferences are single–peaked on the single–dimensional space of choice, Black's (1948) median voter theorem proves that there is a unique outcome under majority rule, and that it coincides with the ideal profile of the voter at the median of the distribution. Proposition 1 proves that single–peakedness is a feature of log–concave preferences. Using log–concavity of preferences, Caplin and Nalebuff (1991) show that if the space of choices is multidimensional, the unique outcome under a 64%–majority rule is the ideal profile of the mean voter. Preservation of unimodality is an interesting result for models of Political Economy because it allows to ensure that politicians's preferences will share the relevant features of voters' preferences. For instance, each $f_i(\cdot)$ may represent the preference of an individual for the provision of a public good net of her individual tax contribution. Thus, $f_0(\cdot)$, the preference of the representative that gets elected with the most votes, shares the same peakedness properties than the electors that voted him. Thus, these results make possible to study how voters' preferences are mapped into political decisions when it is not decided through a referendum but by means of the elected representatives.

4.3 Monotone Likelihood Ratio

Models of moral hazard require that optimal signals used by agents keep a one–to–one correspondence with agents' types [Holmström (1979); Laffont (1989, §11)]. The tools presented in this paper allows to extend this basic model to environments where the resulting distribution of single–dimensional types is the outcome of the combination of several signals of the agents. The idea in this case is that agents use different signals. However, since there is an aggregate type defined either through the addition (10) or the mapping (5), the different signals can also be summarized by the aggregate signal that also keeps a one-to-one correspondence with the aggregate type. Thus, the principal can use individual signals or their aggregate to infer the type of agents.

Preservation of MLR holds for a wider class of functions than the preservation of IHR. The reason is that it relies on properties common to the family of distributions that are TP_2 and not only PF_2. The following Proposition and Corollary prove these basic results.

PROPOSITION 2: *A continuously differentiable probability density function $f(x, \alpha)$ is TP_2 in x_i and the indexing parameter α, if and only if it is MLR.*

PROOF: See Appendix. ▪

COROLLARY 2: *If $f_i(\theta_i, \alpha_i)$, $i = 1, 2$, are MLR and θ_1 and θ_2 are independently distributed, then $f_0(\theta_0, \alpha_0)$ defined according to equation (8) is also MLR.*

PROOF: Proposition 3 ensures that $f_i(\theta_i, \alpha_i)$, $i = 1, 2$, are TP_2 while Corollary 1 ensures that the composition of functions that are TP_2 is also TP_2. Thus, $f_0(\theta_0, \alpha_0)$ is MLR. ▪

5 When is Bundling Optimal?

The multidimensional screening literature has frequently found bundling to be optimal even when types are not correlated. In screening models with a single–dimensional type it is possible to rank the profitability of different mechanisms depending on the hazard rate ordering of different informational structures. A well known sufficient condition to compare the optimal solutions of different mechanisms is to require a particular hazard rate ordering of the involved distributions. Since in the present model different types aggregate into a single dimensional variable, the hazard rate ordering is endogenously explained by the properties of the convolution of distributions of type components. Optimal contracts critically depend on the value of the hazard rate of the corresponding distribution, and I thus have to establish how large is the hazard rate of the convolution distribution $F_0(\theta)$ relative to $F_i(\theta)$. Proposition 3 shows that θ_0 dominates in hazard rate to θ_i if the support of the distribution of types is restricted to $\Re_+$.

PROPOSITION 3: *Let $F_i(\theta_i)$ be IHR, i.e., $r_i'(\theta_i) > 0$ in θ_i on $\{\theta_i > 0 : F_i(\theta_i) < 1\}$, for $i = 1, 2$. Let $F_0(\theta_0)$ denote the convolution distribution of $\theta_0 = \theta_1 + \theta_2$, with hazard rate $r_0(\theta_0)$. Then $r_0(\theta) \leq \min\{r_1(\theta), r_2(\theta)\}$ on $\{\theta > 0 : F_i(\theta) < 1; i = 0, 1, 2\}$.*

PROOF: See Appendix. ■

The result of Proposition 3 implies that the distribution of θ_0 always puts more weight on higher values than the distribution of θ_1 or θ_2 . Therefore, given some value $\theta^\star$, the probability that $\theta_0 > \theta^\star$, $1 - F_0(\theta^\star)$, always exceeds the probability that $\theta_i > \theta^\star$, $1 - F_i(\theta^\star)$. This intuitive result is formalized in the following corollary.

COROLLARY 3: *If $r_0(\theta) \leq r_i(\theta)$ on $\{\theta > 0 : F_i(\theta) < 1; i = 0, 1, 2\}$, then θ_0 first order stochastically dominates θ_i.*

PROOF: Since $r_i(\theta_i) = -d \log[1 - F_i(\theta_i)]/d\theta_i$ it follows that $\forall \theta > 0$:

$$1 - F_0(\theta) = \exp\left[-\int_0^\theta r_0(z)dz\right] \geq \exp\left[-\int_0^\theta r_i(z)dz\right] = 1 - F_i(\theta), \tag{15}$$

and therefore $F_0(\theta) \leq F_i(\theta)$ $\forall \theta > 0$, which is the definition of first order stochastic dominance, of θ over θ_i. ■

Observe that the first order stochastic dominance ordering of stochastic objective functions analyzed by Athey (2000, §2) arises endogenously here within this framework of asymmetric information dealing with multiple characteristic of agents. According to Laffont and Tirole's interpretation (1993, §1.4–1.5), Proposition 3 means that the distribution of θ_0 is more favorable than the distribution of θ_1 or θ_2, *i.e.*, it puts more weight on the high valuation customers. Corollary 3 shows that this result can be generated by a model of individual stochastic demands if the existence of an independent but systematically positive type shock θ_2 ensures that the actual purchase (or valuation) θ_0 is always higher in first order stochastic dominance sense than the expected purchase (or valuation) θ_1.[12]

Addressing the problem of optimal pricing by a monopolist, Maskin and Riley (1984, §4) already considered the effect of changes in the distribution of consumer types on the shape of the nonlinear tariffs. Combining the above Proposition 3 with Proposition 5 of Maskin and Riley (1984), it follows that a nonlinear schedule based on $F_0(\cdot)$ involves higher markups and expected profits than the nonlinear tariff based only on $F_i(\cdot)$ while integrating out the effect of θ_j for all consumption levels.[13]

Observe that these results can only be ensured to hold in models where the support of the distributions are restricted to $\Re_+$.[14] But this is actually the case in many multi-dimensional screening problems discussed before. For instance, in the auction literature, θ_i is the nonnegative value of each individual object to be auctioned. In Palfrey's (1983), the individual valuation of the bundle that defines a multi–object auction cannot be lower than the sum of valuations of the individual components. Thus, the auctioneer is more uncertain about the valuation of bidders when selling a bundle because the addition of the valuations of its components puts more probability weight at the top end of potential bids whenever the distribution of the valuation of each item is IHR. In a discriminatory auction, the mechanism needs to be more powerful in order to induce self–selection of all participating agents, leading to uniformly larger rent extraction for low types.

Dealing with informational decentralization, Baron and Besanko (1992) and Gilbert and Riordan (1995) identify θ_i with the non–negative marginal cost of the members of integrated alliance of suppliers. The role of Assumption 3 in Gilbert and Riordan (1995) is to ensure that the convolution distribution of the marginal cost of the bundle has a smaller hazard rate than the distribution of its components, and thus characterize properly a separating equilibria with several sources of asymmetric information (marginal costs of each firm producing an element of the bundle). However, since these are distributions of on nonnegative variables, Corollary 3 shows that Gilbert and Riordan's assumption is in fact not necessary, because it is ensured by the convolution of IHR distributions of the aggregate type components.

Many other agency problems could define environments where the support of type components is constrained in a natural way. For instance, we could think of $\theta_1 \in \Re_+$ as general skills of workers before being hired (*e.g.*, acquired through education and/or working experience in other jobs). If hired, workers will develop some specific skills and abilities due to learning by doing, and therefore increase their productivity. It is not unreasonable within this framework to exclude the possibility of negative learning, and

[12] Although there is no *a priori* reason to assume that expectations are biased, there is enough evidence not to rule out this possibility. See Miravete (2000a, §5).

[13] In addition to Maskin and Riley (1984, §4), Propositions 3 and 4 of Miravete (2001) proves formally these well known results.

[14] If type components are not restricted to $\Re_+$, the hazard rate dominance has to be imposed exogenously as in the regulatory model of Laffont and Tirole (1993).

thus θ_2 would also be restricted to take only positive values. The principal could then design contracts contingent on either the credentials and qualifications of the worker, or on the actual performance after learning.[15] The previous results show that the principal will prefer to tie workers' compensation to their performance.

In all these models, bundling is optimal even when type components are independently distributed. The result is a consequence of Corollary 3 as well as for the hazard rate of the distribution being inversely related to the optimal markup of the monopolist. When there is more than one source of asymmetry of information it is more difficult to screen consumers of different types. If all sources of asymmetry tend to get every consumer type closer to the highest type possible, the monopolist has to introduce important distortions to reduce the informational rents of infra–marginal types and thus enforce the incentive compatibility of contracts.

Finally, we can address the case of correlated valuations. Bundling will be preferred to independent screening depending on the effect of correlation on the hazard rate ordering of the distribution of the aggregate type relative to its components, –normally a highly nonlinear relation–. For instance, consider the reference case where θ_1 and θ_2 are independent and $F_0(\theta) \leq F_1(\theta)\ \forall \theta \in \Theta \subseteq \Re$. Assume now that type components θ_1 and θ_2 are negatively correlated, and denote by $F_0^\star(\cdot)$ the cumulative distribution of $\theta_0 = \theta_1 + \theta_2$ under negative correlation.[16] The distribution of θ_0 is now less dispersed, with less mass of probability at the tails of the distribution than if θ_1 and θ_2 were independent.[17] In some sense the monopolist is now "less uncertain" about the value that consumer types may take, because there is a larger mass of probability around the mean of θ_0. Thus, for low values of θ_0 (below the mean), the probability of finding a type above a given θ_0 is higher under negative correlation than under independence. Therefore, the hazard rate function is lower under negative correlation than under independence for low values of θ_0. Just the contrary holds for high values of θ_0, *i.e.*, the hazard rate of the distribution with negative correlation will exceed that of the distribution of independent type components. If $r^\star(\theta) \leq r(\theta)$ only for low values of θ, then for high valuation customers bundling markups will be lower under negative correlation of type components than under the assumption of independence as markups and hazard rate of the distribution of types are inversely related. Types are more concentrated around the mean under negative correlation than under independence, and thus it is necessary to introduce important distortions to distinguish among low consumers and preserve the IC property of the mechanism. The opposite will hold if valuations are positively correlated. But then high valuation customers will face higher markups relative to the independence case.

---

[15] A model that shares many of these features in Regulatory Economics is Sappington (1982).

[16] One of the few cases where $F^\star(\theta)$ can be written explicitly is that of $\theta = \theta_1 + \theta_2$ where $(\theta_1, \theta_2) \sim BVN[\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho]$. In this case, $\theta \sim N[\mu = \mu_1 + \mu_2, \sigma^2 = \sigma_1^2 + 2\rho\sigma_1\sigma_2 + \sigma_2^2]$. To illustrate the argument of this paragraph, I computed the hazard rate functions of $\theta_1 + \theta_2$ under independence, $r(\theta)$, and under perfect negative correlation ($\rho = -1$), $r^\star(\theta)$. For the case where $\mu_1 = 0$, $\sigma_1^2 = 1$, $\mu_2 = 1$, $\sigma_2^2 = 0.5$, I found that $r^\star(\theta) > r(\theta)\ \forall \theta > 0.12$.

[17] Note that the effect of correlation fulfills the requirements of the particular kind of mean preserving spread that Courty and Li (2000, §3) need to order their type space defined by probability distributions.

This discussion may shed some light on the discrepancy of predictions between Palfrey (1983) and Armstrong (2000) regarding whether bundling auctions will be optimal for the seller. According to the analysis of the previous paragraph, the presence of positive correlation favors the sale of bundled products in multi–object auctions. This result is opposite to that of Armstrong (2000), but similar to the one of Palfrey (1983) with a small number of bidders. Types in the present model, as well as in Palfrey's (1983) are continuous. The result would hold however if types take a discrete number of possible values, a sequence, and frequency functions replace distributions. The critical difference with Armstrong (2000) model appears then to be that type components aggregate into a single parameter while his model consider a truly multidimensional model although with binary types only.

6 Concluding Remarks

This paper has described some preservation results that may prove useful in the field of mechanism design. Most results, except those related to unimodality, also hold for non–continuously differentiable frequency functions that fulfill the discrete version of *total positivity* [Karlin (1968, §8)]. The most important application to screening is to prove that solutions based on the aggregate type introduce higher distortions to enforce incentive compatibility constraints because it comprises several sources of asymmetric information. This result requires that each type component has an IHR distribution function defined on positive compact support.

In order to solve explicitly the problem of multidimensional screening and show that bundling is an equilibrium feature of these models Armstrong (1996, 2000) suggest the use polar coordinates to ensure that incentive compatibility holds along rays. The present model takes a different approach by modeling the money–value of each type dimension so that their linear aggregation convey some economic meaning to the difference between bundled (centralized, *ex–post*) and unbundled (decentralized, *ex–ante*) screening. By making the different type dimensions to lie on the real line, the mechanism design problem becomes tractable and the focus of the analysis is shifted to the statistical properties of the distributions of types, that now identify whether the bundled solution is preferred.

A limitation of the analysis, common to the literature of multidimensional screening, is the study of cases where types are correlated. Convolutions and compositions are well defined when type components are independent. If they are not independent, the distribution of the aggregate type is no longer the product of the individual distributions of type components, and except in some few cases, the distribution of the aggregate type can only be characterized by numerical methods [Miravete (2001, §4.3)], This confirms the opinion of Armstrong (2000) and Rochet and Choné (1998) to rely on numerical methods to gain some new insights of the features of equilibria in general models of multidimensional screening.

References

ADAMS, W.J., AND J.L. YELLEN (1976): Commodity Bundling and the Burden of Monopoly." *Quarterly Journal of Economics*, 90, 475–498.

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

ARMSTRONG, M. (2000): "Optimal Multi–Object Auctions." *Review of Economic Studies*, 67, 455–481.

ATHEY, S. (2000): "Characterizing Properties of Stochastic Objective Functions." Mimeo, MIT.

AUSUBEL, L. (1991): "The Failure of Competition in the Credit Card Market," *American Economic Review*, 81, 50–81.

AVERY, C. AND T. HENDERSHOTT (2000): "Bundling and Optimal Auctions of Multiple Products." *Review of Economic Studies*, 67, 483–497.

BARON, D.P., AND D. BESANKO (1999): "Informational Alliances." *Review of Economic Studies*, 66, 743–768.

BARON, D.P., AND D. BESANKO (1992): "Information, Control, and Organizational Structure." *Journal of Economics & Management Strategy*, 1, 237–275.

BIAIS, B., D. MARTIMORT, AND J.C. ROCHET (2000): "Competing Mechanisms in a Common Value Environment." *Econometrica*, 68, 799–837.

BLACK, D. (1948): "On the Rationale of Group Decision–Making." *Journal of Political Economy*, 56, 23–34.

CAILLAUD, B., R. GUESNERIE, AND P. REY (1992): "Noisy Observation in Adverse Selection Models." *Review of Economic Studies*, 59, 595–615.

CAPLIN, A. AND B. NALEBUFF (1991): "Aggregation and Social Choice: A Mean Voter Theorem." *Econometrica*, 59, 1–23.

COURTY, P. AND H. LI (2000): "Sequential Screening." *Review of Economic Studies*, 67, 697–717.

GANTMACHER, F.R. (1959): *Matrix Theory*. Chelsea Publishing Company.

GILBERT, R.J. AND M.H. RIORDAN (1995): "Regulating Complementary Products: A Comparative Institutional Analysis." *RAND Journal of Economics*, 26, 243–256.

HOLMSTRÖM, B (1979): "Moral Hazard and Observability," *Bell Journal of Economics*, 10, 74–91.

KARLIN, S. (1968): *Total Positivity*, Vol. I. Stanford University Press.

LAFFONT, J.J. (1989): *The Economics of Uncertainty and Information*. MIT Press.

LAFFONT, J.J. AND J. TIROLE (1986): "Using Cost Observations to Regulate Firms." *Journal of Political Economy*, 94, 614–641.

LAFFONT, J.J. AND J. TIROLE (1993): *A Theory of Incentives in Procurement and Regulation*. MIT Press.

MASKIN, E. AND J. RILEY (1984): "Monopoly with Incomplete Information." *Rand Journal of Economics*, 15, 171–196.

MCAFEE, R.P. AND J. MCMILLAN (1988): "Multidimensional Incentive Compatibility and Mechanism Design." *Journal of Economic Theory*, 46, 335–354.

MCAFEE, R. P., J. MCMILLAN, AND M. WHINSTON (1989): "Multiproduct Monopoly, Commodity Bundling, and the Correlation of Values." *Quarterly Journal of Economics*, 103, 371–383.

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

Miravete, E.J. (2000a): "Choosing the Wrong Calling Plan? Ignorance, Learning, and Risk Aversion." CEPR Discussion Paper No. 2562.

Miravete, E.J. (2000b): "Estimating Demand for Local Telephone Service with Asymmetric Information and Optional Calling Plans." CEPR Discussion Paper No. 2635.

Miravete, E.J. (2001): "Quantity Discounts for Taste–Varying Consumers." CEPR Discussion Paper No. 2699.

Palfrey, T.R. (1983): "Bundling Decisions by a Multiproduct Monopolist with Incomplete Information." *Econometrica*, 51, 463–483.

Panzar, J.C. and D.S. Sibley (1978): "Public Utility Pricing under Risk: The Case of Self–Rationing." *American Economic Review*, 68, 887–895.

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

Riordan, M.H. and D.E.M. Sappington (1987): "Awarding Monopoly Franchises." *American Economic Review*, 77, 375–387.

Sappington, D. (1982): "Optimal Regulation of Research and Development under Imperfect Information." *Bell Journal of Economics*, 13, 354–368.

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

Spence, A.M. (1980): "Multi–Product Quantity–Dependent Prices and Profitability Constraints." *Review of Economic Studies*, 47, 821–842.

Spulber, D.F. (1992): "Optimal Nonlinear Pricing and Contingent Contracts." *International Economic Review*, 33, 747–772.

Wilson, R.B. (1995): "Nonlinear Pricing and Mechanism Design," in H. Amman, D. Kendrick, and J. Rust (eds.): *Handbook of Computational Economics*, Vol. I. North–Holland.

Appendix 1

• Proof of Proposition 1

In order to prove parts (a) and (b) of this Theorem let first study the total positivity properties of the function $\delta : \Re \to \{0, 1\}$ defined as follows:

$$\delta(x-y) = \begin{cases} 0 & \text{if} \quad x < y \\ 1 & \text{otherwise} \end{cases} \qquad (A.1)$$

From Definition 6, $\delta(x-y)$ is PF_2 if $\forall x_1, x_2 \in X \subseteq \Re$ and $\forall y_1, y_2 \in Y \subseteq \Re$, such that $x_1 < x_2$ and $y_1 < y_2$, the following condition holds:

$$\begin{vmatrix} \delta(x_1-y_1) & \delta(x_1-y_2) \\ \delta(x_2-y_1) & \delta(x_2-y_2) \end{vmatrix} \geq 0. \qquad (A.2)$$

Simple analysis of all possible cases show that $\delta(x-y)$ is PF_2. It is then straightforward to show that $\hat{\delta}(x-y) = 1 - \delta(x-y)$ is also PF_2. By Lemma 3, $\hat{\gamma}(\theta_i)$, the convolution of $\hat{\delta}(x-\theta_i)$ and $f_i(\theta_i)$ is PF_2. Hence:

$$\hat{\gamma}(\theta_i) = \int_{\Re} \hat{\delta}(x-\theta_i) f_i(\theta_i) d\theta_i = \int_{-\infty}^{x} f_i(\theta_i) d\theta_i = F_i(\theta_i = x), \qquad (A.3)$$

because $\hat{\delta}(x-\theta_i) = 1$ only if $x < \theta_i$, and therefore the cumulative distribution function $F_i(\theta_i)$ is PF_2. Similarly, $\gamma(\theta_i)$ the convolution of $\delta(x-\theta_i)$ and $f_i(\theta_i)$ is also PF_2, which in this case implies that:

$$\gamma(\theta_i) = \int_{\Re} \delta(x-\theta_i) f_i(\theta_i) d\theta_i = \int_{x}^{\infty} f_i(\theta_i) d\theta_i = \overline{F}_i(\theta_i = x), \qquad (A.4)$$

because $\delta(x-\theta_i) = 1$ only if $x \geq \theta_i$, and the survival function $1 - F_i(\theta_i)$ is also PF_2.

To prove part (c), note that by Definition 2, it follows that the hazard rate is $r_i(\theta_i) = -\overline{F}_i'(\theta_i)/\overline{F}_i(\theta_i)$ on $\{\theta_i \in \Theta_i : F_i(\theta_i) < 1\}$, which has to be increasing in Θ_i because by part (b) of this Theorem, $\overline{F}_i(\theta_i)$ is log–concave, and according to Definition 1, this implies that the quotient $\overline{F}_i'(\theta_i)/\overline{F}_i(\theta_i)$ is decreasing in Θ_i.

Similarly, to prove part (d), note that part (a) of this Theorem ensures that $F_i(\theta_i)$ is log–concave, which by Definition 1 implies that $l_i'(\theta_i) \leq 0$.

In order to prove part (e), observe that since $f_i(\theta)$ is log–concave, for $x_1 < x_2$ Definition 1 requires:

$$\frac{f_i'(x_1)}{f_i(x_1)} \geq \frac{f_i'(x_2)}{f_i(x_2)}, \qquad (A.5)$$

or equivalently:

$$\begin{vmatrix} f_i'(x_1) & f_i(x_1) \\ f_i'(x_2) & f_i(x_2) \end{vmatrix} \geq 0. \tag{A.6}$$

Assume that $\theta_i^\star$ is such that $f_i'(\theta_i^\star) = 0$. If $\theta_i^\star = x_2$, then condition $(A.6)$ implies that $f_i'(x_1) f_i(\theta_i^\star) \geq 0$. Since $f_i(\theta_i^\star) > 0$, it must be the case that $f_i'(x_1) \geq 0$ for $x_1 < \theta_i^\star$. Conversely, if $\theta_i^\star = x_1$, then $-f_i'(x_2) f_i(\theta_i^\star) \geq 0$. Thus, it must be the case that $f_i'(x_2) \leq 0$ for $x_2 > \theta_i^\star$. Therefore, if $\theta_i^\star$ exists, $f_i(\theta_i)$ is increasing for values of $\theta_i < \theta_i^\star$ and decreasing for $\theta_i > \theta_i^\star$. Otherwise, if $\theta_i^\star$ does not exists, $f_i(\theta_i)$ is either monotone increasing or decreasing. Thus, $f_i(\theta_i)$ is unimodal. ■

- **Proof of Proposition 2**

Density function $f(x, \alpha) > 0$ is TP_2 in x and α if for $x_1 < x_2$ and $\alpha_1 < \alpha_2$:

$$D = \begin{vmatrix} f(x_1, \alpha_1) & f(x_1, \alpha_2) \\ f(x_2, \alpha_1) & f(x_2, \alpha_2) \end{vmatrix} \geq 0, \tag{A.7}$$

Assume, without loss of generality, that $\alpha_1 = \alpha$, $\alpha_2 = \alpha + \Delta_\alpha$, with $\Delta_\alpha > 0$. Then, using common properties of determinants it is straightforward to show:

$$D = \begin{vmatrix} f(x_1, \alpha) & f(x_1, \alpha + \Delta_\alpha) \\ f(x_2, \alpha) & f(x_2, \alpha + \Delta_\alpha) \end{vmatrix} = \begin{vmatrix} f(x_1, \alpha) & f(x_1, \alpha + \Delta_\alpha) - f(x_1, \alpha) \\ f(x_2, \alpha) & f(x_2, \alpha + \Delta_\alpha) - f(x_2, \alpha) \end{vmatrix}$$

$$= \begin{vmatrix} f(x_1, \alpha) & \dfrac{f(x_1, \alpha + \Delta_\alpha) - f(x_1, \alpha)}{\Delta_\alpha} \\ f(x_2, \alpha) & \dfrac{f(x_2, \alpha + \Delta_\alpha) - f(x_2, \alpha)}{\Delta_\alpha} \end{vmatrix} \cdot \Delta_\alpha \geq 0. \tag{A.8}$$

Since $\Delta_\alpha > 0$, it follows that:

$$D_\alpha = \lim_{\Delta_\alpha \to 0} \left(\frac{D}{\Delta_\alpha} \right) = \begin{vmatrix} f(x_1, \alpha) & f_\alpha(x_1, \alpha) \\ f(x_2, \alpha) & f_\alpha(x_2, \alpha) \end{vmatrix} \geq 0. \tag{A.9}$$

Proceeding similarly with x and assuming that $x_1 = x$, $x_2 = x + \Delta_x$, with $\Delta_x > 0$, it follows that:

$$D_\alpha = \begin{vmatrix} f(x, \alpha) & f_\alpha(x, \alpha) \\ f(x + \Delta_x, \alpha) & f_\alpha(x + \Delta_x, \alpha) \end{vmatrix}$$

$$= \begin{vmatrix} f(x, \alpha) & f_\alpha(x, \alpha) \\ \dfrac{f(x + \Delta_x, \alpha) - f(x, \alpha)}{\Delta_x} & \dfrac{f_\alpha(x + \Delta_x, \alpha) - f_\alpha(x, \alpha)}{\Delta_x} \end{vmatrix} \cdot \Delta_x \geq 0, \tag{A.10}$$

so that:

$$D_{x\alpha} = \lim_{\Delta_x \to 0} \left(\frac{D_\alpha}{\Delta_x} \right) = \begin{vmatrix} f(x,\alpha) & f_\alpha(x,\alpha) \\ f_x(x,\alpha) & f_{x\alpha}(x,\alpha) \end{vmatrix} \geq 0. \tag{A.11}$$

But observe that:

$$D_{x\alpha} = f^2(x,\alpha) \cdot \frac{\partial^2 \ln f(x,\alpha)}{\partial x \partial \alpha} \geq 0, \tag{A.12}$$

which according to Definition 3 hold if and only if $f(x,\alpha) > 0$ is MLR. ▪

• Proof of Proposition 3

Suppose not, *i.e.*, for instance assume that $r_1(\theta) < r_0(\theta)$:

$$\frac{f_1(\theta)}{\overline{F}_1(\theta)} < \frac{f_0(\theta)}{\overline{F}_0(\theta)}. \tag{A.13}$$

Using the definition of convolution in equation (11), this inequality is equivalent to the following three inequalities:

$$f_1(\theta)\overline{F}_0(\theta) - f_0(\theta)\overline{F}_1(\theta) < 0, \tag{A.14a}$$

$$f_1(\theta) \int_0^\infty \overline{F}_1(\theta - z) f_2(z) dz - \overline{F}_1(\theta) \int_0^\infty f_1(\theta - z) f_2(z) dz < 0, \tag{A.14b}$$

$$\int_0^\infty \left[f_1(\theta)\overline{F}_1(\theta - z) - \overline{F}_1(\theta) f_1(\theta - z) \right] f_2(z) dz < 0. \tag{A.14c}$$

Since $f_2(\theta) \geq 0$ on $0 \leq \theta < \infty$, it must be the case that the term between brackets is negative $\forall \theta \geq 0$. But observe that this condition then requires:

$$\frac{f_1(\theta)}{\overline{F}_1(\theta)} \leq \frac{f_1(\theta - z)}{\overline{F}_1(\theta - z)} \qquad \forall z \geq 0, \tag{A.15}$$

so that $F_1(\theta_1)$ should be decreasing hazard rate. Similarly, $r_2(\theta) < r_0(\theta)$ violates $F_2(\theta_2)$ being IHR. Contradiction. ▪