CARESS Working Paper 01–01

Screening Through Bundling

Eugenio J. Miravete[†⋆]

December 18, 2000

Abstract

This paper studies a class of multidimensional screening problems where different type dimensions lie on the real line, and applies results related to the preservation of log–concavity under convolution. In particular, it is shown 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. The application of these results ensures that screening of agents is feasible even when individual demands are stochastic. The paper discusses several applications in different areas of economics. 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.

:/.../papers/penn10/penn10.tex

1 Introduction

Consider the following motivating example. A consumer wants to buy a service provided by a regulated public utility. In many occasions, the public utility offers a nonlinear tariff to reduce consumers' informational rents by offering 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. Alternatively, the public utility may offer different contract options to consumers who are now required to choose among them before their consumption is realized. These contract options are generally defined by a monthly fee and a particular rate per unit of consumption, or more in general, through a fully nonlinear tariff plus, in some cases, some capacity limits to consumption (*e.g.*, power load) and/or 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 while the second screens consumers sequentially. First, it makes consumers choose among different tariff options based on their expectation of future purchase levels, and later each tariff option introduces different additional discounts or premia on the difference between expected and realized demand. However, the important point is that when consumers sign up for a particular contract option, they do not commit to any given level of consumption. At 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.

This paper studies distributional properties common to a general class of models involving agents with more than one type dimension for which it is possible to characterize well behaved, fully separating equilibria. For most of the analysis, I consider that type dimensions are related as follows:

$$\theta_0 = \theta_1 + \theta_2, \tag{1}$$

where θ_1 and θ_2 are stochastically independent, so that the distribution of θ_0 is the convolution:

$$F_0(\theta_0) = \int_{\Theta_j} F_i(\theta_0 - \theta_j) dF_j(\theta_j), \tag{2}$$

where indices can be reversed because the convolution is a commutative operation.[1] The structure of equations (1) – (2) 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. The following is a non–exhaustive list of models that fit the above structure:

[1] The characteristic function of $F_0(\theta_0)$ is just the product of the Fourier transforms of the distributions of its components [Hirschman and Widder (1955, §2.5); Karlin (1968, §7.1–7.3)].

- *Expected Consumption.* This is the case discussed above. 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 [Ausubel (1991); Miravete (2000b); Miravete (2001)].

- *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 that is easily observable for the principal, such as weather conditions. Thus, the monopolist may solve the optimal state contingent tariff that makes payment and discounts dependent on the realization of such variable [Spulber (1992)] or simply design a tariff that mainly target individual differences although taking into account the effect on individual demands of other variables such as temperature [Panzar and Sibley (1978)] that are not known at the time of subscribing the power capacity option.

- *Regulation.* The possibility of errors in the appraisal of the cost function of the regulated firm allows agencies to establish regulatory contracts based either on realized or expected costs. The literature on the optimality of linear contracts show that these simple contracts are robust to the existence of an additive noise, θ_2 [Caillaud, Guesnerie, and Rey (1992); Laffont and Tirole (1986)].[2]

- *Procurement.* Awarding procurement contracts involves frequently firms bidding when they are uncertain about their future marginal costs [Riordan and Sappington (1987)]. Alternatively, the government could ask for a share of total revenues or profits to the awarded franchisees, thus making transfers a function of actual rather than expected costs.

The setup of equations $(1) - (2)$ 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. For instance:

- *Common Agency.* In Biais, Martimort, and Rochet (2000) θ_1 is the signal of the value of an asset that is privately observed by an agent, while θ_2 is the agent's endowment shock of such risky asset. Both type dimensions aggregate into a single parameter θ_0 representing the marginal valuation of an agent for the asset to be traded.

- *Informational Alliances.* Baron and Besanko (1999), distinguish between unit costs of independent suppliers, θ_1 and θ_2, and the type (cost profile) of an informational

[2] In these models, θ_2 enters linearly in the objective function so that neither the incentive compatibility and participation constraints are changed. This is not necessarily the case in the models of *expected consumption*. In this latter case, the distribution of θ_2 affects the shape of the tariff based on θ_1. See Miravete (2001, §3.1) for a detailed discussion.

alliance, θ_0, that can be formed to contract with the principal by consolidating the private information of agents.[3]

- *Multi–Object Auctions.* In Palfrey (1983), θ_1 and θ_2 are the independent individual valuations of two objects that the principal may auction separately or in a bundle, and for which each potential buyer is willing to bid up to θ_0.[4]

As some of these examples point out, the distinction between θ_1 and θ_2 might be useful if we want to model separately the different elements behind the optimal behavior of agents. 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, we can follow 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 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 or any of its components. 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).

Multidimensional screening addresses the existence of more than a single source of heterogeneity among agents and study how they condition the features of optimal contracts. Existing models have shown not only how easily can they become intractable, but also and more importantly, how dealing with multiple dimensions of agents' types leads in most cases to optimal bunching, non–monotone contracts, and even an optimal positive

[3] Other related papers are Baron and Besanko (1992) and Gilbert and Riordan (1995).

[4] Its dual 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 of two–part tariffs, or bundling all these taste parameters to design a single two–part tariff. The difference is that θ_0 is no longer the sum of θ_1 and θ_2, and thus, compositions should replace convolutions to carry out the analysis.

rent extraction for the highest consumer type [Armstrong (1996); Rochet and Choné (1998); Wilson (1995)]. Another common result is that bundling is generally preferred even when individual valuations of its components are independently distributed. The approach followed in the present work allows to explain this equilibrium feature 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 have to be solved. But perhaps, the most important contribution of this paper is to point out some useful results for a structure that is common to many models in diverse areas of Economics.

The paper is organized as follows. Section 2 presents the mathematical tools needed to prove these preservation results. Mathematical generality is kept to a minimum in order to prove that *total positivity* is preserved under composition and *log–concavity* is preserved under convolution. Section 3 presents the main results of interest for Economics, *i.e.*, that IHR, MLR, and unimodality of the density functions are preserved for appropriately defined problems. Section 3 also proves that IHR is preserved under convolution even if the individual probability density functions are not log–concave. Section 4 discusses Proposition 1 of Biais, Martimort, and Rochet (2000). Section 5 analyzes whether bundling is preferred to independent screening of each type dimension. Section 6 concludes.

2 Totally Positive and Log–Concave Functions

This section presents the minimal mathematical tools needed to prove the preservation results of this paper.

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:

$$F_i(\theta_i) = \int_{\underline{\theta}_i}^{\theta} f_i(z)dz, \tag{3}$$

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:

$$\frac{\partial^2 \log[f_i(\theta_i)]}{\partial \theta_i^2} = \frac{\partial}{\partial \theta_i}\left[\frac{f_i'(\theta)}{f_i(\theta)}\right] \leq 0 \quad \text{on} \quad \Theta_i. \tag{4}$$

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 hold. 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:

$$r_i(\theta_i) = \frac{f_i(\theta_i)}{1 - F_i(\theta_i)} \qquad \text{on} \quad \{\theta_i \in \Theta_i : F_i(\theta_i) < 1\}, \tag{5}$$

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 characterize the type of an agent given an 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.

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:

$$\frac{\partial^2 \ln[f_i(\theta_i, \alpha)]}{\partial\theta_i \partial\alpha} \geq 0. \tag{6}$$

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 $\theta_i^\star \in \arg\max_{\theta_i} f_i(\theta_i)$.

To ease the presentation of the preservation of the MLR result in the next section, it is convenient to assume a more general definition of the "aggregation" of type dimensions than the one described in equation (1). Therefore, assume that:

$$\theta_0 = T(\theta_1, \theta_2) : \Re^2 \to \Re. \tag{7}$$

In general, when we deal with the aggregation of dimensions of agents' own types, equation (1) suffices to fully characterize such aggregation as monotone transformations of utility functions represent the same set of preferences. However, equation (7) 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, transformation (7) may prove useful when dealing with the aggregation of preferences across individuals that carry some sort of weighting.

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 (7). Let define the following composition operation:[5]

$$M(\theta_0, \zeta) = \int\limits_{\Theta_j} K(\theta_0, \theta_j) L(\theta_j, \zeta) dF_j(\theta_j), \tag{8}$$

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 $L(\cdot)$. The nonlinear function (7) 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 (7). 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.

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) 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 $1 \leq m \leq n$:

$$\begin{vmatrix} g(x_1, y_1) & g(x_1, y_2) & \cdots & g(x_1, y_m) \\ g(x_2, y_1) & g(x_2, y_2) & \cdots & g(x_2, y_m) \\ \vdots & \vdots & \ddots & \vdots \\ g(x_m, y_1) & g(x_m, y_2) & \cdots & g(x_m, y_m) \end{vmatrix} \geq 0. \tag{9}$$

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 (8). The following Lemma states this identity[6].

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:*

$$M(\theta_0, \zeta) = \int\limits_{\Theta_2} K(\theta_0, \theta_2) L(\theta_2, \zeta) dF_2(\theta_2) = \int\limits_{\Theta_1} K(\theta_1, \zeta) L(\theta_0, \theta_1) dF_1(\theta_1), \tag{10}$$

is also TP_n.

[5] Kernel $K(\cdot)$ in composition equation (8), 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.

[6] 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)]

Proof: Without loss of generality, let $n = 2$. By definition of TP_2, the composition $M(x, y)$ defined in (10) 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:

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

$$= \int\limits_{z_1 <} \int\limits_{z_2} \begin{vmatrix} K(x_1, z_1) & K(x_1, z_2) \\ K(x_2, z_1) & K(x_2, z_2) \end{vmatrix} \cdot \begin{vmatrix} L(z_1, y_1) & L(z_2, y_1) \\ L(z_1, y_2) & L(z_2, y_2) \end{vmatrix} dF_z(z_1)dF_z(z_2) \geq 0, \tag{11}$$

where the last inequality is the *Basic Composition Formula* that relates compositions of totally positive functions.[7] 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 PF_2. ■

An immediate consequence of the application of the *Basic Composition Formula* is the following result that will be used in Section 4.

Corollary 1: *If $K(x, y)$ is TP_m and $L(x, y)$ is TP_n, then $M(x, y)$, the composition defined in* (10), *is $PF_{\min(m,n)}$.*

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 $(1) - (2)$. Equation (8) reduces to the convolution case of equations $(1)-(2)$ when $M(\theta_0, \zeta) = F_0(\theta_0)$, $K(\theta_0, \theta_j) = F_i(\theta_0 - \theta_j)$, and $L(\theta_j, \zeta) = 1$. 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 $1 \leq m \leq n$:

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

[7] 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).

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:*

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

is also PF_n.

COROLLARY 2: *If $f_1(\theta_1)$ is PF_m and $f_2(\theta_2)$ is PF_n, then $f_0(\theta_0)$, the probability density function defined by convolution* (13)*, is $PF_{\min(m,n)}$.*

3 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 and PF_2, and MLR and TP_2 respectively.

3.1 Increasing Hazard Rate

The mathematical results of the previous section show that the smoothness properties of *Pólya frequency functions* of the same order 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 such 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 inequalities hold:

$$\begin{vmatrix} g(x_1) & g(x_1 - \Delta) \\ g(x_2) & g(x_2 - \Delta) \end{vmatrix} = \Delta \cdot \begin{vmatrix} \dfrac{g(x_1) - [g(x_1 - \Delta)]}{\Delta} & g(x_1 - \Delta) \\ \dfrac{g(x_2) - [g(x_2 - \Delta)]}{\Delta} & g(x_2 - \Delta) \end{vmatrix} \geq 0. \tag{14a}$$

Since $\Delta > 0$, we can take limits in the latter determinant to obtain:

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

leading to:

$$\frac{g'(x_1)}{g(x_1)} \geq \frac{g'(x_2)}{g(x_2)}, \tag{14c}$$

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

We 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 3: *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. ∎

These results are applicable not only to screening problems. For instance, they could also be applied to significant issues in Political Economy since the preservation of single–peakedness of preferences is ensured.[8] But regarding mechanism design, which is the focus of this paper, 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. When individual demands are stochastic this opens the possibility of sequential screening. Thus the principal has the ability to design contracts targeting either the different dimensions of agents' types, or their aggregate. While these two problems are equivalent from an interpretative perspective, the results of Proposition 1 and Corollary 3 ensure that the principal can induce separating equilibria under the two alternative approaches, although nothing implies that they lead to the same solution.[9] However, if types represents valuations of different goods or characteristics, the principal has now the ability to screen each dimension separately or simultaneously through bundling.

In many problems, as those of nonlinear pricing, the critical assumption is the IHR property of the distribution of types and not the more restrictive assumption of log–concavity of the corresponding density functions. The above results imply that if the density functions of type components are log–concave, then the screening problems could also be solved when demands are stochastic, but only at the cost of reducing the set of distributions that could be used for modeling these problems. Fortunately, the set of problems that can be solved is not reduced because, as the next Proposition shows, IHR is preserved under convolution regardless of the log–concavity of the respective density functions.[10]

PROPOSITION 2: *If $F_1(\theta_1)$ and $F_2(\theta_2)$ are IHR, then their convolution $F_0(\theta_0)$ defined in equation* (2) *is also IHR.*

---

[8] 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, $f_i(\cdot)$ 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. It should not be difficult to show using these tools that an equivalent of the median or mean voter theorem will not hold when the votes of the representatives are not weighted by the population that elected them.

[9] See Miravete (2001, §6) for a discussion on the relative profitability of the different screening approaches using information on actual and expected usage of local telephone service.

[10] Observe that Proposition 2 does not exclude the possibility that the convolution of distributions is IHR even when at least one of the convoluting distributions is decreasing hazard rate. See Karlin (1968, 3.8.C) for an example.

PROOF: See Appendix. ■

3.2 Monotone Likelihood Ratio

Models of moral hazard requires that optimal signals used by agents keep a one–to–one relationship with agents' types [Holmström (1979); Laffont (1989, §11)]. The approach of 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. This is the result of Corollary 4 below.

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 3: *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 4: *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. ■

4 Discussion

In a recent work, Biais, Martimort, and Rochet (2000) develop a common agency model in which agents' types have two dimensions that lie on the real line and define a single dimensional aggregate. Thus, according to one of the modeling alternatives described before in the Section 1, these authors make the necessary assumptions on the distribution of this latter variable in order to characterize an equilibrium in nonlinear schedules that depends exclusively on θ_0, thus reducing the dimensionality of the screening problem.

Theorem 1 of Biais, Martimort, and Rochet (2000) argues that such procedure is not restrictive because log–concavity is preserved under convolution. They claim that the convolution of a log–concave density $f_1(\theta_1)$, with any *arbitrary* distribution $f_2(\theta_2)$ leads to probability distribution $F_0(\theta_0)$, and survival functions $1 - F_0(\theta_0)$ that are log–concave for the convolution $\theta_0 = \theta_1 + \theta_2$. Their only requirement is that $f_1(\theta_1)$ and the *arbitrary* density $f_2(\theta_2)$ are defined on a bounded support.

Figures 1 and 2 cast some doubt on the validity of this result. In these figures, the first row represents the probability density functions $f_i(\theta_i)$, $i = 0, 1, 2$. The second row pictures the ratio $f_i'(\theta_i)/f_i(\theta_i)$ to analyze the log–concavity of the density functions, while the third row shows the ratio $l_i(\theta_i) = f_i(\theta_i)/F_i(\theta_i)$ to analyze the log–concavity of the distribution functions. Finally, the bottom row describes the behavior of the hazard rate, $r_i(\theta_i)$.

Figure 1 shows the convolution of a uniform distribution defined on the unit interval, with a beta distribution with parameters $p = 0.4$ and $q = 0.5$, also defined on the unit interval. The third column represents the distribution of their convolution defined on the $[0, 2]$ interval.[11] As it is well known, the uniform is a log–concave distribution with increasing hazard rate. The beta distribution –defined on a bounded support as required by Biais, Martimort, and Rochet (2000)–, may or may not be log–concave depending on the values of the indexing parameters p and q. If these are sufficiently small, the ratio $f_2'(\theta_2)/f_2(\theta_2)$ becomes increasing (log–convex density), the distribution function fails to be log–concave, and the hazard rate includes a region where it decreases. The consequence for the convolution distribution is that small values of p and q make the density of θ_2 sufficiently log–convex to turn the convolution density sufficiently peaked, so that there is a nontrivial region in the neighborhood of the mode of θ_0 where $f_0'(\theta_0)/f_0(\theta_0)$ is increasing, thus violating the log–concavity of $f_0(\theta)$. Although there might be log–concave distribution functions whose densities are not log–concave, the ratio $f_0(\theta_0)/F_0(\theta_0)$ in the figure also rejects such hypothesis around the mode of θ_0 and in a neighborhood of its lower bound. Finally, the bottom figure of the third column clearly shows that the hazard rate is not increasing for the whole support of Θ_0, thus contradicting the presumed log–concavity of the convolution survival function $1 - F_0(\theta_0)$.

Figure 2 removes the restriction of bounded supports for the distributions of θ_1 and θ_2. The first column represents a standard lognormal distribution, and the second column shows a standard normal distribution. The third column is the lognormal–normal convolution. This case has some appeal for economic modeling since the normally distributed variable may represent an error of measurement in the appraisal of each individual's own type which in addition, due to some economic reason, it is restricted to take positive values in many models.[12] It is well known that the lognormal density is not log–concave and that it is characterized by a decreasing hazard rate as θ_1 increases [Sweet (1990)]. Consequently, and according to Corollary 2, the convolution density function is not log–concave. However, its distribution is log–concave [Bagnoli and Bergstrom (1989)], a property that is preserved under convolution. The normal density is log–concave, and thus by Theorem 1, its distribution and survival functions are log–concave, and therefore it is characterized by the IHR property. However, as it can easily be confirmed looking at the third column

[11] Pham and Turkkam (1994) study this type of convolution. A general reference is Johnson, Kotz, and Balakrisnahan (1995, §25.8). To obtain the convolution density function, the range of integration was divided in 10,000 intervals. For each one of these intervals, the convolution was computed using a 40–points Gauss–Legendre quadrature.

[12] Hawkins (1991) studies in detail the basic properties of this convolution. For a general overview, see Johnson, Kotz, and Balakrisnahan (1995, §14.8). Romberg integration was used to compute the convolution density with a minimum of 10,000 divisions of the initial range $\Theta_0 = [-25, 75]$. Convergence required an error of integration smaller than 10^{-8} to define the final range of integration.

of Figure 2, the convolution of a log–concave density as the normal, and an *arbitrary* distribution such as the lognormal, does not ensure that the convolution distribution is IHR in the case of unbounded supports.

The proof of Theorem 1 in Biais, Martimort, and Rochet (2000) ignores that the convolution is a commutative operation, and thus, the distribution $f_2(\theta_2)$ (in my notation) plays no role; it just smears the effect of the endowment shock on the domain Θ_0 according to the rule of the distribution of asset values $f_1(\theta_1)$. Intuitively, the same result should be obtained by spreading the effect of θ_1 according to the distribution of θ_2. Thus, the proof should be true if $f_1(\theta_1)$ is replaced by $f_2(\theta_2)$ and *vice versa*, but then, $f_2(\theta_2)$ cannot be any *arbitrary* density defined on a bounded support, but rather a *log–concave* density without a necessarily bounded support.

The results of the previous sections allow me to prove this point more rigourously. Any single dimensional density function that takes only nonnegative values is, by Definition 6, at least PF_1. This is the case of all convoluting distributions of Figures 1 and 2. However, by Lemma 3, only log–concave densities are PF_2. Thus, the uniform in Figure 1 and the normal distribution in Figure 2 are PF_2. As shown by Corollary 2, the convolution of *Pólya frequency functions* of different order is also a *Pólya frequency function* of order equal to the lower order of the convoluting distributions. Therefore, the convolutions of Figures 1 and 2 are necessarily PF_1, which while still well defined as densities, lack the log–concavity property, a sufficient condition to prove that the convolution distribution and survival functions are both log–concave. It could still be argued that Theorem 1 of Biais, Martimort, and Rochet (2000) does not make any inference about the log–concavity of the convolution density function but only about the distribution and survival function of the convolution. However, both examples in Figure 1 and 2 show that there are regions in Θ_0 where log–concavity of $F_0(\theta_0)$ and/or the IHR property fails to hold.

There are two alternatives to overcome this difficulty. One is to assume that both density functions are log–concave. As log–concavity is preserved under convolution, Corollary 3 ensures that the convolution distribution has the desired properties. Alternatively, only IHR should be assumed to hold for each convoluting distribution, and thus Proposition 2 ensures that the convolution distribution is also IHR regardless of the log–concavity of the involved densities. This latter approach, common in models of single dimensional screening, is less restrictive and ensures the existence of separating equilibria in models that target agent's aggregate informational types.

5 When is Bundling Optimal?

The multidimensional screening literature has frequently found bundling to be optimal unless types are strongly positively 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 find and/or require a particular hazard rate ordering of the involved distributions. Since in the model of this paper, different types aggregate into a single dimensional variable, it is worth exploring whether the convolution of type components define a hazard rate ordering of the distribution of θ_0 relative to θ_1 and θ_2. Since optimal contracts critically depend on the value of the hazard rate of the corresponding distribution I have to establish how large is the hazard rate of the convolution distribution $F_0(\theta_0)$ relative to those of the components of the *ex–post* type. Proposition 4 shows that θ_0 dominates in hazard rate to θ_i if the support of the distributions is restricted to $\Re_+$.

PROPOSITION 4: *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 $\theta = \theta_1 + \theta_2$, 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 4 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 $\hat{\theta}$, the probability that $\theta_0 > \hat{\theta}$ always exceeds the probability that $\theta_i > \hat{\theta}$. This intuitive result is formalized in the following corollary.

COROLLARY 5: *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. ■

The first order stochastic dominance ordering of stochastic objective functions analyzed by Athey (2000, §2) arises naturally 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 4 means that the distribution of θ_0 is more favorable than the distribution of θ_1 or θ_2. Corollary 5 shows that this result could be generated endogenously in 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 stochastic sense than the expected purchase (or valuation) θ_1. Although there is no *a priori* reason to assume that expectations are biased, there is enough evidence not to rule this possibility [Miravete (2000a, §5)].

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 4 with Proposition 5 of

Maskin and Riley (1984), it follows that a nonlinear schedule based on $F_0(\theta_0)$ involves higher markups than the nonlinear tariff based only on $F_i(\theta_i)$ while integrating out the effect of θ_j for all consumption levels.

Observe that these results can only be ensured to hold in models where the support of the distributions are restricted to $\Re_+$. Otherwise the hazard rate dominance has to be imposed exogenously. But this actually the case in many multidimensional screening problems. For instance, in the auction literature, θ_i is nonnegative the value of each individual object to be auctioned. 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.[13]

It is also plausible that many agency problems we can 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 could 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 thus θ_2 could 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.[14]

In all these models bundling is optimal even when type components are independently distributed. The result is a consequence of Corollary 5 as well as for the hazard rate of the distribution being inversely related to the optimal markup of the monopolist. When there are 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 induce self–selection.

Finally, the model also identifies why bundling is optimal unless types are strongly positively correlated.

For instance, consider the reference case where θ_1 and θ_2 are independent and $F_0(\theta_0) \leq F_1(\theta\_0)\ \forall\theta \in \Theta \subseteq \Re$. 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.[15] 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. In some

[13] The role of Assumption 3 in Gilbert and Riordan (1995) is to ensure that the convolution distribution has a smaller hazard rate than the distribution of its components. Since these distributions are defined on nonnegative variables, Corollary 5 shows that this assumption is not necessary, because it is implied by the convolution of IHR distributions.

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

[15] 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, $r(\theta)$,

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. Thus, 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 large customers *ex–post* nonlinear pricing 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 θ are inversely related. Consumer 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. Thus, the results of this paper in general need to be qualified for particular cases where type components are allowed to be correlated.

6 Concluding Remarks

The application of convolution results were introduced in Economics by Karlin (1959), although they had never been explicitly used in economic modeling. 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 Definition 5 [Karlin (1968, §8)].

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

---

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 $r^\star(\theta) > r(\theta)\ \forall\theta > 0.12$.

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

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.

Bagnoli, M. and T. Bergstrom (1989): "Log–Concave Probability and Its Applications." Working Paper 89–23, University of Michigan.

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.

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.

Hawkins, D.M. (1991): "The Convolution of the Normal and Lognormal Distributions." *South African Statistical Journal*, 25, 99–128.

Hirschman, I.I. and D.V. Widder (1955): *The Convolution Transform*. Princeton University Press.

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

Johnson, N.L., S. Kotz, and N. Balakrishnan (1995): *Continuous Univariate Distributions*, 2nd edition. John Wiley & Sons.

Karlin, S. (1959): *Mathematical Methods and Theory in Games, Programming, and Economics*, Vol. II. Addison–Wesley.

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.

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. ####.

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.

Pham, T.G. and N. Turkkan (1994): "Reliability of a Standby System with Beta–Distributed Component Lives." *IEEE Transactions on Reliability*, 43, 71–75.

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.

Sweet, A.L. (1990): "On the Hazard Rate of The Lognormal Distribution." *IEEE Transactions on Reliability*, 39, 325–328.

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

Zwillinger, D (1992): *Handbook of Integration*. Jones and Barlett Publishers.

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:

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

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:

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

There are six possible cases:

1. If $x_1 < x_2 < y_1 < y_2$, then $0 \cdot 0 - 0 \cdot 0 = 0$,
2. If $x_1 < y_1 \leq x_2 < y_2$, then $0 \cdot 0 - 1 \cdot 0 = 0$,
3. If $x_1 < y_1 < y_2 \leq x_2$, then $0 \cdot 1 - 1 \cdot 0 = 0$,
4. If $y_1 \leq x_1 < y_2 \leq x_2$, then $1 \cdot 1 - 1 \cdot 0 = 1$,
5. If $y_1 \leq x_1 < x_2 < y_2$, then $1 \cdot 0 - 1 \cdot 0 = 0$,
6. If $y_1 < y_2 \leq x_1 < x_2$, then $1 \cdot 1 - 1 \cdot 1 = 0$.

Thus $\delta(x-y)$ 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:

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

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:

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

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 ensures that $l'_i(\theta_i) \leq 0$.

Finally, to prove part (e) assume, without loss of generality, that $x_1 < x_2$ and $0 = y_1 < y_2 = \Delta$. Then, from the definition of PF_2 and making use of common properties of determinants it is ensured that:

$$\begin{vmatrix} f_i(x_1) & f_i(x_1-\Delta) \\ \\ f_i(x_2) & f_i(x_2-\Delta) \end{vmatrix} = \begin{vmatrix} f_i(x_1)-f_i(x_1-\Delta) & f_i(x_1-\Delta) \\ \\ f_i(x_2)-f_i(x_2-\Delta) & f_i(x_2-\Delta) \end{vmatrix}$$

$$= \begin{vmatrix} \dfrac{f_i(x_1)-f_i(x_1-\Delta)}{\Delta} & f_i(x_1-\Delta) \\ \dfrac{f_i(x_2)-f_i(x_2-\Delta)}{\Delta} & f_i(x_2-\Delta) \end{vmatrix} \cdot \Delta \geq 0. \qquad (A.5)$$

Since $\Delta > 0$, we can take limits in the latter determinant to obtain:

$$\lim_{\Delta\to 0} \begin{vmatrix} \dfrac{f_i(x_1)-f_i(x_1-\Delta)}{\Delta} & f_i(x_1-\Delta) \\ \dfrac{f_i(x_2)-f_i(x_2-\Delta)}{\Delta} & f_i(x_2-\Delta) \end{vmatrix} = \begin{vmatrix} f'_i(x_1) & f_i(x_1) \\ \\ f'_i(x_2) & f_i(x_2) \end{vmatrix} \geq 0. \qquad (A.6)$$

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

Note that by Definition 2, parts (b) and (c) of Theorem 1 are equivalent. Thus, I have to prove that the survival function of the convolution distribution is log–concave, *i.e.*, for $x_1 < x_2$ and $y_1 < y_2$:

$$\begin{vmatrix} 1-F_0(x_1-y_1) & 1-F_0(x_1-y_2) \\ 1-F_0(x_2-y_1) & 1-F_0(x_2-y_2) \end{vmatrix} =$$

$$\begin{vmatrix} \int [1-F_1(x_1-z)]f_2(z-y_1)dz & \int [1-F_1(x_1-z)]f_2(z-y_2)dz \\ \\ \int [1-F_1(x_2-z)]f_2(z-y_1)dz & \int [1-F_1(x_2-z)]f_2(z-y_2)dz \end{vmatrix} =$$

$$\begin{vmatrix} \int [1-F_1(x_1-z)]f_2(z-y_1)dz & \int f_1(x_1-z)[1-F_2(z-y_2)]dz \\ \\ \int [1-F_1(x_2-z)]f_2(z-y_1)dz & \int f_1(x_2-z)[1-F_2(z-y_2)]dz \end{vmatrix} =$$

$$\int\limits_{z_1<}\int\limits_{z_2} \begin{vmatrix} 1-F_1(x_1-z_1) & f_1(x_1-z_2) \\ 1-F_1(x_2-z_1) & f_1(x_2-z_2) \end{vmatrix} \cdot \begin{vmatrix} f_2(z_1-y_1) & 1-F_2(z_2-y_1) \\ f_2(z_1-y_2) & 1-F_2(z_2-y_2) \end{vmatrix} dz_1 dz_2 \geq 0. \quad (A.7)$$

The second determinant just states the survival function $1 - F_0(\cdot)$ in terms of the distributions $F_1(\cdot)$ and $F_2(\cdot)$. The third determinant makes use of Assumption 1 to ensure the existence of the density functions and integrates the expressions in the second column of the third determinant by parts using the convolution identity:

$$\int F_1(x-z)f_2(z-y)dz = \int f_1(x-z)F_2(z-y)dz, \qquad (A.8)$$

while the double integral is the *Basic Composition Formula*. Observe that for that last expression to be positive and thus ensure that the distribution $F_0(\cdot)$ is IHR, each of the determinants has to be positive. Assuming without loss of generality that $0 = z_1 < z_2 = \Delta$, this condition requires that the first determinant is positive:

$$[1-F_1(x_1)]f_1(x_2-\Delta) - [1-F_1(x_2)]f_1(x_1-\Delta) \geq 0, \qquad (A.9)$$

which implies:

$$\frac{f_1(x_2-\Delta)}{1-F_1(x_2-\Delta)} \cdot \frac{1-F_1(x_2-\Delta)}{1-F_1(x_2)} \geq \frac{f_1(x_1-\Delta)}{1-F_1(x_1-\Delta)} \cdot \frac{1-F_1(x_1-\Delta)}{1-F_1(x_1)}. \qquad (A.10)$$

But since $\Delta > 0$ and $x_1 < x_2$:

$$\frac{f_1(x_2-\Delta)}{1-F_1(x_2-\Delta)} \geq \frac{f_1(x_1-\Delta)}{1-F_1(x_1-\Delta)}, \qquad (A.11)$$

which is just the hypothesis that $F_1(\cdot)$ is IHR. Similarly, comparing the other terms of inequality $(A.10)$:

$$\frac{1-F_1(x_2-\Delta)}{1-F_1(x_2)} \geq \frac{1-F_1(x_1-\Delta)}{1-F_1(x_1)}, \qquad (A.12)$$

which is equivalent to:

$$\begin{vmatrix} 1-F_1(x_1) & 1-F_1(x_1-\Delta) \\ \\ 1-F_1(x_2) & 1-F_1(x_2-\Delta) \end{vmatrix} \geq 0. \qquad (A.13)$$

This is the condition for the survival function $1 - F_1(\cdot)$ to be log–concave, which holds by assumption as $F_1(\cdot)$ is IHR. A similar argument proves that if $F_2(\cdot)$ is IHR, the second determinant in the last inequality of $(A.7)$ is also positive. Thus, $F_0(\cdot)$ is IHR. ∎

• Proof of Proposition 3

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

$$D = \begin{vmatrix} f(x_1,\alpha_1) & f(x_1,\alpha_2) \\ f(x_2,\alpha_1) & f(x_2,\alpha_2) \end{vmatrix} \geq 0, \qquad (A.14)$$

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:

$$\begin{aligned} D &= \begin{vmatrix} f(x_1,\alpha) & f(x_1,\alpha+\Delta_\alpha) \\ f(x_2,\alpha) & f(x_2,\alpha+\Delta_\alpha) \end{vmatrix} \\ &= \begin{vmatrix} f(x_1,\alpha) & f(x_1,\alpha+\Delta_\alpha) - f(x_1,\alpha) \\ f(x_2,\alpha) & f(x_2,\alpha+\Delta_\alpha) - f(x_2,\alpha) \end{vmatrix} \\ &= \begin{vmatrix} f(x_1,\alpha) & \dfrac{f(x_1,\alpha+\Delta_\alpha) - f(x_1,\alpha)}{\Delta_\alpha} \\ f(x_2,\alpha) & \dfrac{f(x_2,\alpha+\Delta_\alpha) - f(x_2,\alpha)}{\Delta_\alpha} \end{vmatrix} \cdot \Delta_\alpha \geq 0. \end{aligned} \qquad (A.15)$$

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

$$D_\alpha = \lim_{\Delta_\alpha \to 0} \left(\frac{D}{\Delta_\alpha} \right) = \begin{vmatrix} f(x_1,\alpha) & f_\alpha(x_1,\alpha) \\ f(x_2,\alpha) & f_\alpha(x_2,\alpha) \end{vmatrix} \geq 0. \qquad (A.16)$$

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

$$\begin{aligned} D_\alpha &= \begin{vmatrix} f(x,\alpha) & f_\alpha(x,\alpha) \\ f(x+\Delta_x,\alpha) & f_\alpha(x+\Delta_x,\alpha) \end{vmatrix} \\ &= \begin{vmatrix} f(x,\alpha) & f_\alpha(x,\alpha) \\ \dfrac{f(x+\Delta_x,\alpha) - f(x,\alpha)}{\Delta_x} & \dfrac{f_\alpha(x+\Delta_x,\alpha) - f_\alpha(x,\alpha)}{\Delta_x} \end{vmatrix} \cdot \Delta_x \geq 0, \end{aligned} \qquad (A.17)$$

so that:

$$D_{x\alpha} = \lim_{\Delta_x \to 0} \left(\frac{D_\alpha}{\Delta_x} \right) = \begin{vmatrix} f(x,\alpha) & f_\alpha(x,\alpha) \\ f_x(x,\alpha) & f_{x\alpha}(x,\alpha) \end{vmatrix} \geq 0. \qquad (A.18)$$

But observe that:

$$D_{x\alpha} = f^2(x,\alpha) \cdot \frac{\partial^2 \ln f(x,\alpha)}{\partial x \partial \alpha} \geq 0, \qquad (A.19)$$

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

• **Proof of Proposition 4**

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

$$\frac{f_1(\theta)}{\overline{F}_1(\theta)} < \frac{f_0(\theta)}{\overline{F}_0(\theta)}. \tag{A.20}$$

Using the definition of convolution in equation (2), this inequality is equivalent to the following three inequalities:

$$f_1(\theta)\overline{F}_0(\theta) - f_0(\theta)\overline{F}_1(\theta) < 0, \tag{A.21a}$$

$$f_1(\theta)\int_0^\infty \overline{F}_1(\theta - z)f_2(z)dz - \overline{F}_1(\theta)\int_0^\infty f_1(\theta - z)f_2(z)dz < 0, \tag{A.21b}$$

$$\int_0^\infty \left[f_1(\theta)\overline{F}_1(\theta - z) - \overline{F}_1(\theta)f_1(\theta - z)\right]f_2(z)dz < 0. \tag{A.21c}$$

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:

$$\frac{f_1(\theta)}{\overline{F}_1(\theta)} \leq \frac{f_1(\theta - z)}{\overline{F}_1(\theta - z)} \qquad \forall z \geq 0, \tag{A.22}$$

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. ∎

Figure 1. Uniform–Beta Convolution

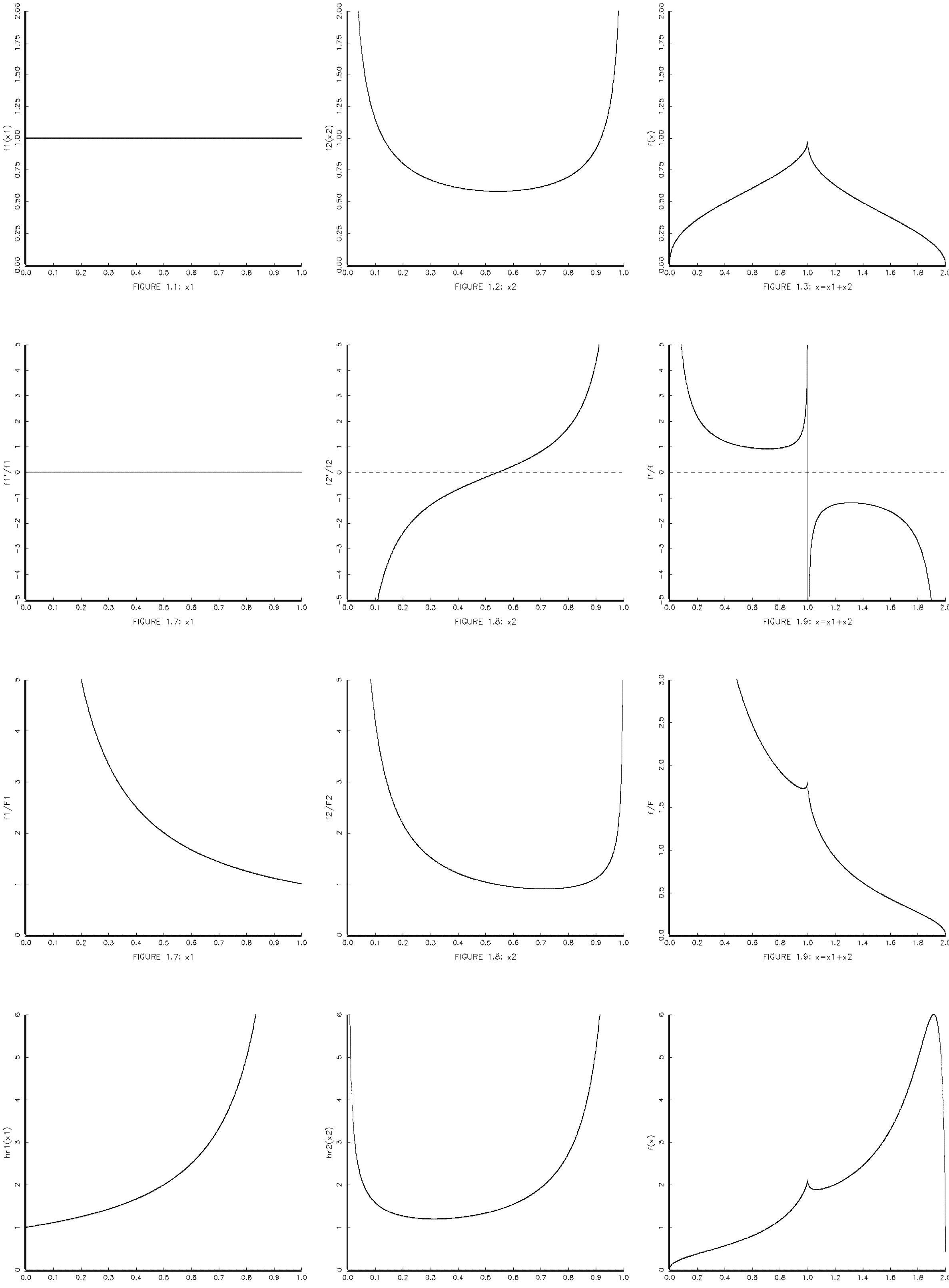

Figure 2. Lognormal–Normal Convolution