Picking Winners in Rounds of Elimination[1]

Suzanne Scotchmer
Dept of Economics and Law School, U.C., Berkeley, and NBER

Junjie Zhou
Department of Mathematics, University of California, Berkeley

Before citing, check the authors' websites for a more recent version.

this version August 12, 2011

[1]We thank the Toulouse Network on Information Technology and the Sloan Foundation for financial support, and Chris Shannon for helpful discussion.

Abstract

We study the optimal way to select projects or agents in environments where information arrives in well defined rounds. Examples include academic environments where review periods are set by policy, aptitude tests such as those given by software developers to programmers applying for jobs, venture capital protocols where the rounds of funding may be stopped before the project is complete, and FDA testing, where drugs can be dropped at well defined junctures. Sequential rounds of elimination reduce the cost of selection, but also reduce the average quality of surviving projects. We characterize the nature of the optimal screening process with and without "memory."

JEL Classifications: C,L2

Keywords: log supermodularity, rounds of elimination, picking winners

1 Introduction

We study economic environments where a principal must select projects or agents from a pool, but cannot observe a candidate's ability or the project's intrinsic worth. A technique for solving this problem is to cast a wide net, and then to eliminate agents or projects that do not perform well. For example, this is how professors are hired. A department initially hires a large pool of assistant professors, gives them a few years to demonstrate their worth, and then makes an up-or-out evaluation. At the full professor stage, the survivors are evaluated again. Those who are not promoted typically leave.

There are many other selection arenas that use a similar technique. Venture capitalists may give early funding to many young start-up firms, but cut them loose ruthlessly when they fail to perform. Drug testing is similar. After the first round of testing, many drugs are dropped, while others go on to another round.

In this paper we ask how such rounds of elimination should be structured. Should there be many rounds or only a few? How should the structure depend on cost? Should standards become tougher or more lax in later rounds? How should the selection at a later round incorporate information generated at earlier rounds?

Different versions of this problem call for different stylizations. In the stylization here, there are natural periods of time in which agents or projects generate signals. The arrival of independent signals in successive periods leads naturally to rounds of elimination. This structure arises naturally in academic life, where evaluation periods are established by policy. Hiring contests are also structured this way. It is common for a software engineering firm to test its applicants by giving them problems to solve. Candidates must pass all the rounds of elimination in order to be considered. How should the rounds of elimination and the use of the signal be structured?

To isolate the issues, we consider only two periods, with two signals (x_1, x_2), independently drawn from a distribution determined by an unobservable ability parameter, μ. The principal wants to select for high values of μ. He can select in a single round of elimination, waiting until the end of period 2 and using both signals, or he can select in double rounds of elimination, using x_1 to winnow the candidates at round one, and then using x_2 to winnow them further. We assume that in both schemes, he is constrained to end up with the same number of survivors.

We distinguish between selection schemes with memory and those without, as in Scotchmer (2008). In a selection scheme with memory, the selection at round two can use the signal generated at round one as well as the signal generated at round two. In a selection scheme without memory, selection at each round can only depend on the signal generated in that round. Sports tournaments are typically selection schemes without memory, whereas promotion in the academic hierarchy typically has memory.

Some of our conclusions are obvious, or at least very intuitive, once stated. However, even the "obvious" conclusions are not always true. They require conditions on the distributions, which we illuminate in this paper. We give a general characterization of optimal selection sets, but also illuminate their special structure when probability densities are log-supermodular. Our work is related to an earlier literature on optimization that uses supermodularity for the conclusion that optimal control variables move monotonically with underlying parameters (Topkis, 1978, Milgrom and Weber, 1982, Milgrom and Shannon, 1995, Athey, 2002). We use log-supermodularity to characterize optimal selection sets rather than optimal control variables.

Our main conclusions are:

- Conditional on a given number of ultimate survivors, the average ability of survivors in a single round of elimination is higher than in any double round of elimination, and the average ability of survivors is higher if the selection scheme has memory than if not.

- A higher cost of holding on to candidates should lead to more stringent screening at round one, less stringent screening at round two, and lower average ability among ultimate survivors.

- In double rounds of elimination without memory (depending on a hazard rate condition), the selection standard should be tougher in round one than in round two.

- If it is optimal to use a sufficient statistic for selection in a single round of elimination, then it is optimal to use the same sufficient statistic in the ultimate round of double elimination, even though the sample has been winnowed in round one, using partial information.

The last point, which follows from the factorization theorem, is perhaps the least intuitive. One might have thought that, in double rounds of elimination, extra weight should be given to x_2 at the second round. The signal x_1 was already used for selection at the first round, and conditional on survival at round one, the signal x_1 is likely to exhibit some "good luck bias." Nevertheless, the same sufficient statistic should be used at the second round as if no prior selection had taken place, although with a less stringent screening standard. No extra weight should be given to x_2, even if there has already been screening on x_1.

In section 2, we describe a simple model. In section 3 we characterize how the selection set should be chosen for a single round of elimination, and record some well known features of probability distributions that lead to monotonic selection criteria. We also give examples, showing that the parameter of interest, μ, may be interpreted in many ways, such as the mean of a distribution, a measure of upside or downside risk, or expected waiting time for an arrival. In section 4 we discuss double rounds of elimination with memory, and in section 5 we discuss double rounds of elimination without memory.

Since we conclude that all the information should be used at every round, section 5 is mainly of interest because, contrary to this conclusion, many selection schemes ignore or de-emphasize information from earlier rounds. Our work implies that, to explain this, one must look elsewhere than simple screening. For example, moral hazard could be a justification. De-emphasizing earlier success maintains an incentive to work harder in later rounds. We intentionally put aside moral hazard problems, because our objective is to isolate the screening problem. Elements of our characterizations will remain when screening and moral hazard are combined.

2 The model

We assume that agents (or projects) are endowed with an underlying parameter μ, which is unobservable. The value of μ is therefore a random variable from the perspective of an observer. The observer wants to select a given number of agents or projects in a way that maximizes the expected value of μ. The underlying parameter μ could be, for example, the profitability of a project, the potential market size of a new innovation, the ability of an assistant professor, the assistant professor's upside potential, or the rate at which the assistant professor thinks of good ideas. The prior distribution on μ is given by a density function h.

The agents (or projects) generate signals $x \in \mathbf{R}^2$ (more generally, $x \in \mathbf{R}^n$) where the draws are assumed to be independent conditional on the underlying value of μ. Each x_i has probability distribution $F(\cdot, \mu)$. Throughout we maintain the assumption that the distribution of signals is atomless with density $f(\cdot, \mu)$.

This simple structure with $x \in \mathbf{R}^2$ permits two rounds of elimination. Agents or projects might be eliminated at the first round, based on x_1, or at the second round, based on (x_1, x_2). A *single round of elimination* means that all agents are kept in the pool until the end of the second period, and selection uses the information generated in both periods, (x_1, x_2). *Double rounds of elimination* mean that some of the agents are eliminated after round one, using only the information x_1. At the second round, selection can take place using both (x_1, x_2) or only x_2. This is the distinction between rounds of elimination *with memory* and *without memory.*

We will use the following notation when it is convenient and not ambiguous:[1] When we write an integral sign without delimiters, we mean the integral on the full support.

$$\begin{aligned} p(x_1, x_2, \mu) &= f(x_1, \mu) f(x_2, \mu) h(\mu) \\ p(x_1, x_2) &= \int p(x_1, x_2, \mu) d\mu \\ p(x_1) &= \int\int p(x_1, x_2, \mu) dx_2 d\mu = \int f(x_1, \mu) h(\mu) d\mu \\ p(x_1, \Delta_2) &= \int_{\Delta_2} p(x_1, x_2) dx_2 \\ p(\Delta_1, x_2) &= \int_{\Delta_1} p(x_1, x_2) dx_1 \end{aligned}$$

Conditional probabilities will also be expressed in this notation:

$$\begin{aligned} p(\mu | x_1, x_2) &= p(x_1, x_2, \mu) / p(x_1, x_2) \\ p(x_2 | x_1) &= p(x_1, x_2) / p(x_1) \\ p(\Delta_2 | x_1) &= p(x_1, \Delta_2) / p(x_1) \\ p(x_2 | x_1, \Delta_2) &= p(x_2 | x_1) / p(\Delta_2 | x_1) \\ p(x_1 | \Delta_1, x_2) &= p(x_1 | x_2) / p(\Delta_1 | x_2) \end{aligned}$$

We will also use sufficient statistics. Using the factorization theorem, a function $\sigma : \mathbf{R}^2 \to \mathbf{R}$ of the signals is *sufficient for* μ if the joint density $p(x_1, x_2, \mu)$ can be written as

$$p(x_1, x_2, \mu) = q(x_1, x_2)\theta(\sigma(x_1, x_2), \mu) \tag{1}$$

for functions $q, \sigma : \mathbf{R}^2 \to \mathbf{R}$ and $\theta : \mathbf{R}^2 \to \mathbf{R}$.

3 Single Round of Elimination

We first consider a single round of elimination, using the information generated in both periods.

In a single round of elimination, the *selection set* is a subset Δ of $\mathbf{R}^2$ such that only the agents or projects that generate signals $(x_1, x_2) \in \Delta$ are chosen. Others are thrown away. When integrated over a set of signals $\Delta \in \mathbf{R}^2$, the total number of survivors is

$$S^s(\Delta) =: \int_\Delta \int p(x_1, x_2, \mu) d\mu dx_1 dx_2$$

[1] We abuse notation to avoid awkwardness, hopefully without confusion. Instead of writing, for example, $p_{(\mu, X_1, X_2)}(\cdot)$ and $p_{(X_1|X_2)}(\cdot)$, as the names of the density functions, we simply write $p(\mu, x_1, x_2)$ and $p(x_1|x_2)$. That is, we write the same thing to refer to the functions themselves as well as to point values of the functions. The context will indicate which interpretation is appropriate.

and their total ability is

$$V^s(\Delta) =: \int_\Delta \int \mu p(x_1, x_2, \mu) d\mu dx_1 dx_2$$

We pose the optimization problem with a constraint on the probability of survival (or number of survivors), $S^s(\Delta) = \Lambda$, $\Lambda \in (0,1)$. Reducing the set Δ leads to fewer survivors. Our problem is to choose the selection set such that the expected ability of survivors is maximized.

The problem we wish to solve is

$$\max_{\Delta} \ V^s(\Delta) \text{ subject to } S^s(\Delta) = \Lambda \tag{2}$$

Following is a general characterization of the solution.

Theorem 1 (Single round of elimination: the optimal selection set) *There exists a finite number α such that, for every solution Δ to (2),*

$$\begin{aligned} E(\mu|x_1, x_2) &\geq \alpha \quad \textit{for a.e.} \quad x \in \Delta \\ E(\mu|x_1, x_2) &\leq \alpha \quad \textit{for a.e.} \quad x \in \mathbf{R}^2 \backslash \Delta \end{aligned} \tag{3}$$

If $\hat{\Delta}$ and Δ are two solutions to (2), then

$$E(\mu|x_1, x_2) = \alpha \textit{ for a.e. } x \in (\hat{\Delta} \backslash \Delta) \cup (\Delta \backslash \hat{\Delta}) \tag{4}$$

Proof: It will be convenient to state the optimization problem using a function g defined as the expected value of μ, given (x_1, x_2).

$$g(x_1, x_2) = \int \mu \ \frac{p(x_1, x_2, \mu)}{\int p(x_1, x_2, \mu) \, d\mu} d\mu = E(\mu|x_1, x_2)$$

Let $d\nu = \left[\int p(x_1, x_2, \mu) \, d\mu\right] dA$ on the interior of its support in $\mathbf{R}^2$. Because $d\nu$ and dA are absolutely continuous with respect to each other, a set that has measure zero with respect to Lebesgue measure also has measure zero with respect to ν measure.

The problem in a single round of elimination can be stated as

$$\max_{\Delta} \int_{\Delta} g\, d\nu \quad \text{subject to:} \quad \int_{\Delta} 1 \cdot d\nu = \nu(\Delta) = \Lambda \tag{5}$$

In appendix A,we first show that for each Λ, there exists a set Δ with the property that:

1. $\nu(\Delta) = \Lambda$,
2. There exists a finite number α such that $g \geq \alpha$ on Δ, and $g \leq \alpha$ on the complement.

We now show that this set is optimal. In particular, for any set A with $\nu(A) = \Lambda$, it must be the case that:

$$\int_A g d\nu \leq \int_{\Delta} g d\nu$$

The first observation is that:

$$\int_{A \backslash \Delta} g d\nu \leq \int_{A \backslash \Delta} \alpha d\nu = \nu(A \backslash \Delta)\alpha = \nu(\Delta \backslash A)\alpha = \int_{\Delta \backslash A} \alpha d\nu \leq \int_{\Delta \backslash A} g d\nu \tag{6}$$

These inequalities hold because $g \leq \alpha$ on $A \backslash \Delta \subset \mathbf{R}^2 \backslash \Delta$, $\alpha \leq g$ on $\Delta \backslash A \subset \Delta$, and $\nu(\Delta \backslash A) = \nu(\Delta) - \nu(A \cap \Delta) = \nu(A) - \nu(A \cap \Delta) = \nu(A \backslash \Delta)$. Thus,

$$\int_A g d\nu = \int_{A \cap \Delta} g d\nu + \int_{A \backslash \Delta} g d\nu \leq \int_{A \cap \Delta} g d\nu + \int_{\Delta \backslash A} g d\nu = \int_{\Delta} g d\nu$$

Take the set A to be another solution, $\hat{\Delta}$, in the expression (6). Then the inequalities are equalities. Because $g \leq \alpha$ on the set $\hat{\Delta} \backslash \Delta$, but the integrals are equal, this implies that $g = \alpha$ almost everywhere on $\hat{\Delta} \backslash \Delta$. Similarly, $g = \alpha$ almost everywhere on $\Delta \backslash \hat{\Delta}$. Hence, $g = \alpha$ almost everywhere on $(\hat{\Delta} \backslash \Delta) \cup (\Delta \backslash \hat{\Delta})$. This shows (3) and (4). ■

Theorem 1 implies that a solution Δ is coupled with a value α that represents the expected ability of the marginal agent. However, without further assumptions, the value α that accompanies Δ is not necessarily unique. When α is not unique, the largest such value is of particular interest, because it represents the infimum of $E\left(\mu | x_1, x_2\right)$ on subsets of selected signals (agents) that have positive measure.

For any solution Δ coupled with a particular α, we can get the other solutions by replacing the part of Δ where $E(\mu|x_1, x_2) = \alpha$ with another set of the same measure where $E(\mu|x_1, x_2) = \alpha$. However, when $\nu(x \in \mathbf{R}^2 : E(\mu|x) = \alpha) = 0$, the optimal selection set is almost unique. By this we mean that if Δ is a solution coupled with α and and $\hat{\Delta}$ is also a solution, it is coupled with the same α, and either that $\nu((\hat{\Delta}\backslash\Delta) \cup (\Delta\backslash\hat{\Delta})) = 0$ or $\nu(\Delta\backslash\hat{\Delta})) = \nu(\hat{\Delta}\backslash\Delta)) = 0$. Every optimal solution is just Δ except on a zero measure set, which is impossible to detect using integration.

When there is a sufficient statistic for μ, Theorem 1 can be restated using the sufficient statistic. For each value $\bar{\sigma} \in \mathbf{R}$, define

$$\bar{E}(\mu|\bar{\sigma}) = \int \mu \frac{\theta(\bar{\sigma}, \mu)}{\int \theta(\bar{\sigma}, \mu) d\mu} d\mu$$

Then it is easy to show that $E(\mu|x_1, x_2)$ has the same value for every signal (x_1, x_2) in the set $\{(x_1, x_2) | \sigma(x_1, x_2) = \bar{\sigma}\}$, and

$$\bar{E}(\mu|\sigma(x_1, x_2)) = E(\mu|x_1, x_2) \tag{7}$$

Theorem 1 thus implies

Corollary 1 *Suppose that σ is a sufficient statistic for μ. There exists a finite number α such that, for every solution Δ to (2),*

$$\bar{E}(\mu|\sigma(x_1, x_2)) \geq \alpha \quad \textit{for a.e.} \quad x \in \Delta$$

$$\bar{E}(\mu|\sigma(x_1, x_2)) \leq \alpha \quad \textit{for a.e.} \quad x \in \mathbf{R}^2\backslash\Delta$$

If $\hat{\Delta}$ and Δ are two solutions to (2), then

$$\bar{E}(\mu|\sigma(x_1, x_2)) = \alpha \textit{ for a.e. } x \in (\hat{\Delta}\backslash\Delta) \cup (\Delta\backslash\hat{\Delta}) \tag{8}$$

3.1 Single Round: Promotion thresholds and monotonicity

The characterization in Theorem 1 and Corollary 1 is too general to be useful. For example, it does not say that if the signal $(\hat{x}_1, \hat{x}_2)$ is larger than some signal (x_1, x_2) in the selection set, then the larger signal $(\hat{x}_1, \hat{x}_2)$ should also be selected. And it does not say that if the value of the sufficient statistic $\sigma(\hat{x}_1, \hat{x}_2)$ is larger than a value $\sigma(x_1, x_2)$ that would be selected, then the larger value of the sufficient statistic should also be selected. These are intuitive conclusions; if they do not hold, then the signal (x_1, x_2) has no natural interpretation.

We will use the mathematical structure of supermodularity. This structure is used widely in economics, following Topkis (1978), Milgrom and Weber (1982), Milgrom and Shannon (1995), and Athey (2002). The literature is concerned with monotone comparative statics: If an optimand is supermodular with respect to the appropriate variables, then the optimizer is a monotonic function of the underlying parameters. Our application here is concerned with optimal selection sets for random variables rather than with optimal control variables. Monotonicity leads to the conclusion that selection sets can be characterized by threshold values.

For convenience, we state the monotonicity assumptions on the density function p. However, they follow from the same properties of f.

The density p satisfies the *monotone likelihood ratio property if*

$$\begin{aligned}
\frac{p(x_1', x_2, \mu')}{p(x_1, x_2, \mu')} &\geq \frac{p(x_1', x_2, \mu)}{p(x_1, x_2, \mu)} \text{ whenever } x_1' > x_1, \mu' > \mu \\
\frac{p(x_1, x_2', \mu')}{p(x_1, x_2, \mu')} &\geq \frac{p(x_1, x_2', \mu)}{p(x_1, x_2, \mu)} \text{ whenever } x_2' > x_2, \mu' > \mu \\
\frac{p(x_1', x_2', \mu)}{p(x_1', x_2, \mu)} &\geq \frac{p(x_1, x_2', \mu)}{p(x_1, x_2, \mu)} \text{ whenever } x_1' > x_1, x_2' > x_2
\end{aligned}$$

Assuming that p is twice differentiable, we will say that the density p is *log supermodular (strictly log supermodular)* if the cross partial derivative of $\log p$ in any of $(x_1, \mu), (x_2, \mu), (x_1, x_2)$ is nonnegative (positive). This is not the general definition, but is equivalent to the general definition when p is differentiable Topkis (1978, p.310). See the papers cited above for the definition and underlying mathematical structure. We will also refer to log supermodularity of θ in (1).

We will also use first-order stochastic dominance. Let two distributions F and G have common supports in $\mathbf{R}$. We say that F *first-order stochastically dominates* G if $F(t) \leq G(t)$ for all t in the supports.

An important fact is that, if F first order dominates G, the expected value of the random variable, or an increasing function of the random variable, is larger when distributed as F than when distributed as G. We will use a slight extension of this fact, stated in the following lemma.

Lemma 1 *Let F and G be two distributions on supports contained in $\mathbf{R}$ such that F first-order dominates G. Suppose that two functions $u, v : \mathbf{R} \to \mathbf{R}$ have the properties that (1) both are nondecreasing and (2) $u(t) \geq v(t)$ for all t. Then:*

$$\int u(t)dF(t) \geq \int v(t)dG(t)$$

Proof:

$$\begin{aligned} & \int u(t)dF(t) - \int v(t)dG(t) \\ = \quad & \int \underbrace{(u(t) - v(t))}_{\geq 0} dF(t) + \left\{ \int v(t)dF(t) - \int v(t)dG(t) \right\} \geq 0 \end{aligned}$$

The first term is nonnegative because the integrand is nonnegative. Nonnegativity of the second term follows because F first-order dominates G and because v is nondecreasing. ■

The following remark records some relationships among the definitions that are used heavily below. The first two bullet points reflect the fact that log-supermodularity is preserved by integration. This property underlies the analysis of Athey (2002), who extended monotone comparative statics to problems where the optimand is the expected value of a log-supermodular function.

Remark 1 *Suppose that $p(x_1, x_2, \mu)$ is log supermodular. Then*

- *Any marginal density function derived from p is log supermodular. For example, $\int p(x_1, x_2, \mu)\, d\mu$ is log supermodular in (x_1, x_2).*

- *Any conditional density function derived from p, such as $p(x_1, \mu | x_2)$, is log supermodular.*

- *Any conditional density function derived from p satisfies the monotone likelihood ratio property.*

- *If $x_1 \geq \hat{x}_1$ and $x_2 \geq \hat{x}_2$, $p(\mu | x_1, x_2)$ first-order stochastically dominates $p(\mu | \hat{x}_1, \hat{x}_2)$.*

- *$E(\mu | x_1, x_2)$ is increasing in x_1 and x_2.*

When we assume that p is log supermodular below, we are adopting all the properties in the Remark.

We also use the following lemma, which links monotonicity properties of a sufficient statistic to log supermodularity of the underlying density function.

Lemma 2 *If p is strictly log supermodular and twice differentiable, and can be written as (1), then σ and θ can be chosen such that θ is strictly log supermodular and σ is increasing in its arguments.*

Proof of Lemma: We want to show that, as long as there is any pair of functions σ and θ such that (1) holds, then there is a pair of function σ and θ such that $\frac{\partial^2}{\partial\sigma\partial\mu} \log \theta(\sigma, \mu) > 0$. Write, for $i = 1, 2$,

$$\begin{aligned}
\frac{\partial^2}{\partial x_1 \partial \mu} \log p(x_1, x_2, \mu) &= \frac{\partial^2}{\partial x_1 \partial \mu} \log g(x_1, x_2) + \frac{\partial^2}{\partial \sigma \partial \mu} \log \theta(\sigma(x_1, x_2), \mu) \frac{\partial}{\partial x_1} \sigma(x_1, x_2) \\
&= \frac{\partial^2}{\partial \sigma \partial \mu} \log \theta(\sigma(x_1, x_2), \mu) \frac{\partial}{\partial x_1} \sigma(x_1, x_2) \\
\frac{\partial^2}{\partial x_2 \partial \mu} \log p(x_1, x_2, \mu) &= \frac{\partial^2}{\partial \sigma \partial \mu} \log \theta(\sigma(x_1, x_2), \mu) \frac{\partial}{\partial x_2} \sigma(x_1, x_2)
\end{aligned}$$

Because $\frac{\partial^2}{\partial x_i \partial \mu} \log p(x_1, x_2, \mu) > 0$, $\frac{\partial}{\partial x_i} \sigma(x_1, x_2) \neq 0$ for $i = 1, 2$. If $\frac{\partial}{\partial x_1} \sigma(x_1, x_2) > 0$, then $\frac{\partial^2}{\partial \sigma \partial \mu} \log \theta(\sigma(x_1, x_2), \mu) > 0$, as required, and $\frac{\partial}{\partial x_2} \sigma(x_1, x_2) > 0$. If $\frac{\partial}{\partial x_1} \sigma(x_1, x_2) < 0$, then $\frac{\partial^2}{\partial\sigma\partial\mu}\log\theta\left(\sigma\left(x_1,x_2\right),\mu\right)<0$ and $\frac{\partial}{\partial x_2}\sigma\left(x_1,x_2\right)<0$. In that case, define $\tilde{\sigma}=-\sigma$. Then $\tilde{\sigma}$ is also sufficient for μ, and $\theta\left(\tilde{\sigma},\mu\right)$ is log supermodular. ■

The following theorem shows that, if there is a sufficient statistic, the optimal selection set can be written as a threshold value on that statistic, that is, $\left\{\left(x_1,x_2\right):\sigma\left(x_1,x_2\right)\geq\bar{\sigma}\right\}$. If the selection scheme has memory, the optimal selection set cannot generally be written, for example, as $\left\{\left(x_1,x_2\right):x_2\geq k_1,x_2\geq k_2\right\}$.

Theorem 2 (Single round of elimination and a sufficient statistic) *Suppose that* p *is log supermodular, and let* Δ *be a selection set that solves* (2). *Suppose that* σ *is a sufficient statistic for* μ. *Then there exists* $\bar{\sigma}$ *such that*

$$\begin{aligned}\sigma\left(x_1,x_2\right) &\geq \bar{\sigma} \text{ for a.e. } x\in\Delta\\ \sigma\left(x_1,x_2\right) &\leq \bar{\sigma} \text{ for a.e. } x\in\mathbf{R}^2\backslash\Delta\end{aligned}$$

Proof: This follows from Theorem 1, and from (7). With log supermodularity, σ is increasing in its arguments, as is $\bar{E}\left(\mu|\bar{\sigma}\right)$. ■

In the next section, we give some examples to show the different meanings that the parameter μ can take. The most familiar case is where μ is the distribution mean, and the sufficient statistic is the mean of the sample. We also consider examples where μ is an extreme point of the support, interpreted as a measure of upside risk or downside risk.

3.2 Examples

Normal Distribution. Let x be a single random draw from a normal distribution with unknown mean μ and known variance v. Then the distribution of x conditional on μ is

$$f(x,\mu)=\frac{1}{\sqrt{2\pi v^2}}e^{-\frac{(x-\mu)^2}{2v^2}},\ x\in\mathbf{R}$$

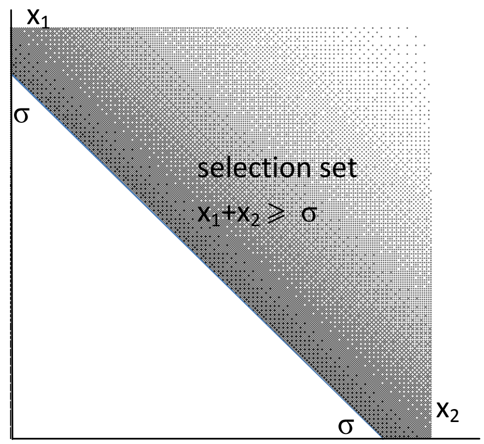

Figure 1: Selection on the mean

One can check directly that f is log supermodular. When a sample $(x_1, x_2, ..., x_n)$ is available, the joint distribution conditional on μ is

$$\begin{aligned}\Pi_{i=1}^{n} f(x_i,\mu) &= \left(\frac{1}{2\pi v^2}\right)^{n/2} e^{-\frac{\sum(x_i-\mu)^2}{2v^2}} = \left(\frac{1}{2\pi v^2}\right)^{n/2} \exp\left\{-\frac{\sum {x_i}^2 - 2\mu\sum x_i + \mu^2}{2v^2}\right\} \\ &= \left(\frac{1}{2\pi v^2}\right)^{n/2} \exp\left\{-\frac{\sum {x_i}^2}{2v^2}\right\} \exp\left\{\frac{2\mu n\bar{x}}{2v^2}\right\} \exp\left\{-\frac{\mu^2}{2v^2}\right\}\end{aligned}$$

Using the factorization theorem, the sample mean, $\bar{x}$, is a sufficient statistic for μ. Applying Theorem 2, the lower bound of the optimal selection set is an affine line with slope -1, that is $\Delta = \{(x_1, x_2)|x_1 + x_2 \geq \sigma\}$ for an appropriate value of the sufficient statistic, σ. This is shown in figure 1.

General exponential distributions. In the exponential family, the distribution of a single random draw, x, conditional on a parameter μ, can be expressed as

$$f(x,\mu) = h(x)\exp(\eta(\mu)\sigma(x) \;-\; A(\mu)),\;\; x \in \mathbf{R}$$

Then, using the factorization theorem for sufficient statistics, σ is a sufficient statistic for μ. Provided $\eta'(\mu)\sigma'(x) > 0$ (both are increasing or both are decreasing), the density function is log supermodular.

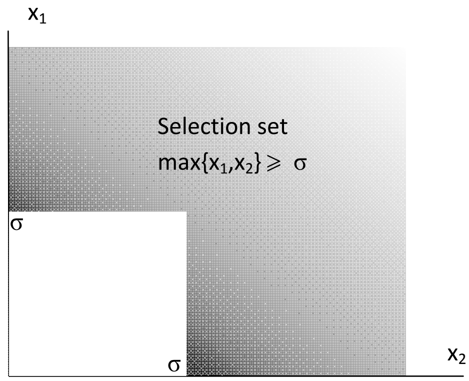

Figure 2: Selection for maximizing upside risk

When there are two random draws, the joint density is $f(x_1, \mu)\, f(x_2, \mu)$, and the sufficient statistic is $\sigma(x_1) + \sigma(x_2)$.

Waiting Time: The random variable x with the following density represents a waiting time, conditional on μ :

$$f(x, \mu) = \frac{e^{-x/\mu}}{\mu}, \mu > 0, x \in \mathbf{R}_+$$

The waiting time itself, x, is a sufficient statistic. If two waiting times are measured, the sufficient statistic is their sum.

Maximizing the upside risk. Suppose x_1 and x_2 are independent and uniformly distributed on the interval $[0, \mu]$. The density of each random draw x is

$$f(x, \mu) = \frac{1}{\mu} 1_{\{0 \leq x \leq \mu\}}(x, \mu)$$

where $1_{\{0 \leq x \leq \mu\}}$ is the indicator function. This density function is log supermodular in (x, μ), because both $1_{\{0 \leq x \leq \mu\}}$ [2] and $\frac{1}{\mu}$ are logsupermodular in (x, μ), and the product of nonnegative log supmodular functions is log supermodular. Therefore $E(\mu|x_1, x_2)$ is weakly increasing in x_1 and x_2.

[2] The set $\{(x, \mu) : 0 \leq x \leq \mu\}$ is a sublattice in $\mathbf{R}^2$. Hence the indicator function $1_{\{0 \leq x \leq \mu\}}(x, \mu)$ is log-supermodular in (x, μ) . See Athey (2002, Lemma 3).

Let

$$\sigma(x_1, x_2) = \max(x_1, x_2)$$

The probability density of (x_1, x_2) can be written as follows:

$$f(x_1, \mu) f(x_2, \mu) = \frac{1}{\mu} 1_{\{0 \leq x_1 \leq \mu\}} \frac{1}{\mu} 1_{\{0 \leq x_2 \leq \mu\}} = \left(1_{\{0 \leq \min(x_1, x_2)\}}\right) \left(\frac{1}{\mu^2} 1_{\{\max(x_1, x_2) \leq \mu\}}\right)$$

Therefore, using the factorization theorem, σ is sufficient for μ. Further, there is an unbiased estimator of μ, β_1, that increases with σ. For this estimator,

$$E(\mu | x_1, x_2) = \beta_1(\max(x_1, x_2))$$

For some number α, an upper contour set of $E(\mu | x_1, x_2)$, hence the selection set Δ, takes the following form:

$$\Delta = \{(x_1, x_2) | \sigma(x_1, x_2) \geq \alpha\} = \{(x_1, x_2) | \max(x_1, x_2) \geq \alpha\}$$

Minimizing the downside risk. Now suppose x_1 and x_2 are independent and uniformly distributed on the interval $[\mu, 1]$. The density of a single random draw, x, is

$$f(x, \mu) = \frac{1}{1 - \mu} 1_{\{\mu \leq x \leq 1\}} (x, \mu)$$

where $1_{\{\mu \leq x \leq 1\}}$ is the indicator function. This density function is log supermodular in (x, μ), because both $1_{\{\mu \leq x \leq 1\}}$ and $\frac{1}{1-\mu}$ are log supermodular in (x, μ), and the product of nonnegative log supmodular functions is log supermodular. Therefore $E(\mu | x_1, x_2)$ is weakly increasing in x_1 and x_2.

Let

$$\sigma(x_1, x_2) = \min(x_1, x_2)$$

The probability density of (x_1, x_2) can be written as follows:

$$f(x_1, \mu) f(x_2, \mu) = \frac{1}{1 - \mu} 1_{\{\mu \leq x_1 \leq 1\}} \frac{1}{1 - \mu} 1_{\{\mu \leq x_2 \leq 1\}} = \left(1_{\{0 \leq \min(x_1, x_2)\}}\right) \left(\frac{1}{(1 - \mu)^2} 1_{\{\min(x_1, x_2) \geq \mu\}}\right)$$

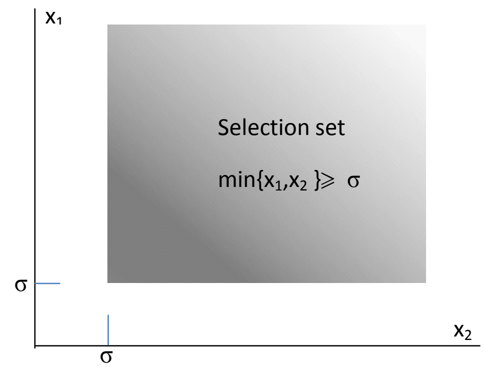

Figure 3: Selection for minimizing downside risk

Therefore, σ is sufficient for μ and there is an unbiased estimator of μ, β_2, that increases with σ. For this estimator,

$$E(\mu|x_1, x_2) = \beta_2(\min(x_1, x_2))$$

For some number α, an upper contour set of $E(\mu|x_1, x_2)$, and therefore the selection set Δ, takes the following form:

$$\Delta = \{(x_1, x_2)| \min(x_1, x_2) \geq \alpha\}$$

4 Double Rounds of elimination with memory

We now suppose that it is costly to collect information in each round, for example, because assistant professors must be paid. In the previous sections, we implicitly assumed that the selection will be made only after two draws. If drawing samples is not costly, this is optimal. However, when sampling is costly, money can be saved by discarding some unpromising projects or agents after the first round. That is, it is optimal to have double rounds of elimination. The potential penalty for saving money in this way is that the first round of elimination might exclude good projects that would be revealed as such if kept for the second round.

We assume that the cost of experimenting is the same in each round, namely, c. This is without loss of generality, assuming it is efficient to begin the experimentation process at all. Since the cost in the first round must be sunk in order to proceed, the (relevant) objective function depends only on the cost in the second round.

The selection process *has memory* if the selection criterion at round two can depend on the signal generated at round one. We view the selection problem as the choice of $\Delta_1 \subset \mathbf{R}$ and $\Delta_2 : \Delta_1 \to \mathcal{A}$, where $\mathcal{A}$ is the set of measurable subsets of $\mathbf{R}$ and where, for each $x_1 \in \Delta_1$, $\Delta_2(x_1) \in \mathcal{A}$ is understood as the selection set at the second round.

The objective is to maximize the expected μ among agents who survive both rounds, minus the cost that must be paid in the second round for survivors of the first round. The number of survivors at the end of the second round is constrained to be Λ. We write the objective function as

$$\begin{aligned} V(\Delta_1, \Delta_2; c) \equiv & \int \int_{\Delta_1} \int_{\Delta_2(x_1)} \mu p(x_1, x_2, \mu)\, dx_2 dx_1 d\mu \\ & -c \int \int_{\Delta_1} \int \mu p(x_1, x_2, \mu)\, dx_2 dx_1 d\mu \end{aligned}$$

We write the number of survivors of both rounds as

$$S(\Delta_1, \Delta_2) =: \int \int_{\Delta_1} \int_{\Delta_2(x_1)} p(x_1, x_2, \mu)\, dx_2 dx_1 d\mu$$

Then the objective is

$$\text{maximize } V(\Delta_1, \Delta_2; c) \text{ subject to } S(\Delta_1, \Delta_2) = \Lambda \tag{9}$$

Theorem 3 (Double elimination with memory: the optimal policy) *Let Δ_1, Δ_2 be the selection sets that solve (9). Then there exists a number λ such that*

$$E(\mu | x_1, \Delta_2(x_1)) - \frac{c}{p(\Delta_2(x_1) | x_1)} - \lambda \geq 0 \text{ for a.e. } x_1 \in \Delta_1 \tag{10}$$

$$E(\mu | x_1, \Delta_2(x_1)) - \frac{c}{p(\Delta_2(x_1) | x_1)} - \lambda \geq 0 \text{ for a.e. } x_1 \notin \Delta_1$$

$$\text{for a.e. } x_1 \in \Delta_1, \begin{cases} E(\mu | x_1, x_2) - \lambda \geq 0 \text{ for a.e. } x_2 \in \Delta_2(x_1) \\ E(\mu | x_1, x_2) - \lambda \leq 0 \text{ for a.e. } x_2 \notin \Delta_2(x_1) \end{cases} \tag{11}$$

Proof: In order to characterize the solution, it is convenient to reformulate the objective as a Lagrange function where the choice variables are indicator values $I(x_1) \in \{0,1\}$ for each $x_1 \in \mathbf{R}$ and $J(x_1, x_2) \in \{0,1\}$ for each $(x_1,x_2) \in \mathbf{R} \times \mathbf{R}$ such that $I(x_1) = 1$ There is a one-to-one relationship between the indicator functions I, J and the selection sets Δ_1, Δ_2, given by

$$\Delta_1 = \{x_1 \in \mathbf{R} : I(x_1) = 1\}$$

$$\Delta_2(x_1) = \{x_2 \in \mathbf{R} : J(x_1, x_2) = 1\} \text{ for each } x_1 \in \Delta_1$$

We will sometimes refer to the optimum as the optimal indicator functions I, J and sometimes as the optimal selection sets Δ_1, Δ_2.

The Lagrange function to be maximized is

$$\begin{aligned}
&\mathcal{L}(I, J) = \int\int\int I(x_1) J(x_1, x_2) \quad \mu \quad p(x_1, x_2, \mu) \, dx_2 dx_1 d\mu \\
&-c \int\int\int I(x_1) \quad p(x_1, x_2, \mu) \, dx_2 dx_1 d\mu \\
&-\lambda \left[\int\int\int I(x_1) J(x_1, x_2) \quad p(x_1, x_2, \mu) \, dx_2 dx_1 d\mu - \Lambda \right]
\end{aligned}$$

The first-order conditions are $S(\Delta_1, \Delta_2) = \Lambda$ and for each (x_1, x_2),

$$\begin{cases} \dfrac{\partial}{\partial I(x_1)} \mathcal{L}(I, J) \geq 0 & \text{if} \quad I(x_1) = 1 \\ \\ \dfrac{\partial}{\partial I(x_1)} \mathcal{L}(I, J) \leq 0 & \text{if} \quad I(x_1) = 0 \end{cases} \tag{12}$$

$$\begin{cases} \dfrac{\partial}{\partial J(x_1, x_2)} \mathcal{L}(I, J) \geq 0 & \text{if} \quad J(x_1, x_2) = 1 \\ \\ \dfrac{\partial}{\partial J(x_1, x_2)} \mathcal{L}(I, J) \leq 0 & \text{if} \quad J(x_1, x_2) = 0 \end{cases} \tag{13}$$

To interpret these conditions, we write out the values of the partial derivatives.

$$\frac{\partial}{\partial I(x_1)}\mathcal{L}(I,J) = \int\int J(x_1,x_2)\,\mu p(x_1,x_2,\mu)\,dx_2 d\mu - c\int\int p(x_1,x_2,\mu)\,dx_2 d\mu \\ - \lambda\int\int J(x_1,x_2)\,p(x_1,x_2,\mu)\,dx_2 d\mu \quad (14)$$

Given a pair $(x_1,\Delta)\in\mathbf{R}\times\mathcal{A}$, we use the notation

$$E(\mu|x_1,\Delta) =: \int_\Delta \frac{p(x_2|x_1)}{p(\Delta|x_1)}E(\mu|x_1,x_2)\,dx_2 \quad (15)$$

Then

$$\begin{aligned}\frac{\partial}{\partial I(x_1)}\mathcal{L}(I,J) &= \int J(x_1,x_2)\,p(x_1,x_2)\int \mu\frac{p(x_1,x_2,\mu)}{p(x_1,x_2)}d\mu dx_2 \\ &\quad -cp(x_1) - \lambda\int J(x_1,x_2)\,p(x_1,x_2)\,dx_2 \\ &= \int J(x_1,x_2)\,p(x_1,x_2)\,E(\mu|x_1,x_2)\,dx_2 - cp(x_1) - \lambda p(x_1,\Delta_2(x_1))\end{aligned}$$

$$\begin{aligned}&\frac{\partial}{\partial I(x_1)}\mathcal{L}(I,J)\times\frac{1}{p(x_1,\Delta_2(x_1))} \\ &= \int J(x_1,x_2)\frac{p(x_1,x_2)}{p(x_1,\Delta_2(x_1))}E(\mu|x_1,x_2)\,dx_2 - c\frac{p(x_1)}{p(x_1,\Delta_2(x_1))} - \lambda \\ &= \int_{\Delta_2(x_1)}\frac{p(x_2|x_1)}{p(\Delta_2(x_1)|x_1)}E(\mu|x_1,x_2)\,dx_2 - c\frac{1}{p(\Delta_2(x_1)|x_1)} - \lambda \\ &= E(\mu|x_1,\Delta_2(x_1)) - \frac{c}{p(\Delta_2(x_1)|x_1)} - \lambda \qquad (16)\end{aligned}$$

$$\begin{aligned}\frac{\partial}{\partial J(x_1,x_2)}\mathcal{L}(I,J) &= I(x_1)\left[p(x_1,x_2)\int_0^\infty \mu\frac{p(x_1,x_2,\mu)}{p(x_1,x_2)}d\mu - \lambda p(x_1,x_2)\right] \\ &= I(x_1)\,p(x_1,x_2)\times[E(\mu|x_1,x_2)-\lambda] \qquad (17)\end{aligned}$$

The conclusions in the theorem follow from (12) and (13), and (16) and (17). ■

The condition (10) for selection at the first round takes account of the cost that will be incurred in the second round. The value of keeping the agent after round one is diminished by the expected cost. The cost is wasted if the agent will be eliminated later. Averaged over the agents who survive both rounds, the per-agent cost of including the signal x_1 at the first round is $\frac{c}{p(\Delta_2(x_1)|x_1)}$.

At the second round of elimination, the decision maker can use both sample points (x_1, x_2) to select the ultimate survivors. Since both sample points contain information, the selection process should clearly use both. However, the selection is only among agents who survived round one – many sample points are "missing," and the ones that are missing were selected in a systematic way. This means that the conditional distribution of x_2, given the selection in round one, is different than the distribution of x_1 in round one, and different than the distribution of x_2 if all agents were in the sample. Given that there will be a "good luck bias" among the agents who survive round one, one might think that x_2 should be given some special weight in the evaluation at round two.

To put these questions more precisely,

- Should the two sample points (x_1, x_2) be treated symmetrically at the end of round two?
- For the case that there is a sufficient statistic for a single round of elimination, should the selection at the end of round two be based on the same statistic?

Perhaps surprisingly, the following corollary answers these questions affirmatively.

Corollary 2 (Double elimination with memory and a sufficient statistic) *Suppose σ is a sufficient statistic for μ. Let Δ_1, Δ_2 be the selection sets that solve (9). Then for a suitable constant λ,*

$$E\left(\mu|x_1, \Delta_2\left(x_1\right)\right) - \frac{c}{p(\Delta_2(x_1)|x_1)} - \lambda \geq 0 \textit{ for a.e. } x_1 \in \Delta_1$$

$$E\left(\mu|x_1, \Delta_2\left(x_1\right)\right) - \frac{c}{p(\Delta_2(x_1)|x_1)} - \lambda \leq 0 \textit{ for a.e. } x_1 \notin \Delta_1$$

$$\textit{for a.e. } x_1 \in \Delta_1, \begin{cases} \bar{E}\left(\mu|\sigma\left(x_1, x_2\right)\right) \geq \lambda \textit{ for a.e. } x_2 \in \Delta_2\left(x_1\right) \\ \bar{E}\left(\mu|\sigma\left(x_1, x_2\right)\right) \leq \lambda \textit{ for a.e. } x_2 \notin \Delta_2\left(x_1\right) \end{cases}$$

Proof: The characterization of Δ_1 is the same as in Theorem 3, and the characterization of Δ_2 relies on the sufficient statistic instead of (x_1, x_2), using (7). ■

Thus, if there is a sufficient statistic for μ, the selection criterion at the second round depends only on this statistic. It is the same sufficient statistic as is used in a single round of elimination, even though the distributions are different. For example, if it is optimal to use only the sample mean for selection in a single round of elimination, then it is optimal to use the mean at round two – in particular, to weight the signals of the two periods equally – even when a prior selection has been made at round one, based only on x_1. No extra weight should be given to x_2 to compensate for the fact that x_1 has already been used at round one.

As with a single round of elimination, the characterization of the optimum is more useful if the density functions are log supermodular, and therefore satisfy the monotone likelihood ratio property. We now turn to this case.

4.1 Double round with memory: monotonicity and threshold values

With additional structure on the distributions, we can say more about how the signals (x_1, x_2) should be used in two rounds of elimination. In particular, the optimal selection sets (Δ_1, Δ_2) are threshold policies.

Theorem 4 shows that log supermodularity again leads to the conclusion that the optimal selection policy is a threshold policy in each round. Theorem 5 refines this result, showing that selection in the first round should use a threshold for the first signal, and if there is a sufficient statistic, should use a threshold for the sufficient statisic in the second round. Both these threshold results use log supermodularity.

Theorem 4 (Double elimination with memory: promotion using threshold values) *Suppose the distribution p is log supermodular. Given a cost c, let Δ_1, Δ_2 be the selection sets that solve (9). Then the optimal selection sets can be written with threshold values*

$k_1(c), \{k_2(x_1,c) : x_1 \geq k_1(c)\}$ *such that*

$$x_1 \geq k_1(c) \text{ for a.e. } x_1 \in \Delta_1$$

$$x_1 \leq k_1(c) \text{ for a.e. } x_1 \notin \Delta_1$$

$$\text{for a.e. } x_1 \in \Delta_1, \begin{cases} x_2 \geq k_2(x_1,c) \text{ for a.e. } x_2 \in \Delta_2(x_1) \\ x_2 \leq k_2(x_1,c) \text{ for a.e. } x_2 \notin \Delta_2(x_1) \end{cases}$$

Proof: Taking Δ_2 first, log supermodularity implies that $E(\mu|x_1,x_2)$ is increasing in x_2. Therefore the conclusion follows from Theorem 3.

For Δ_1, referring to the proof of Theorem 3, rewrite the derivative (14) of the Lagrange function as

$$\begin{aligned} \frac{\partial}{\partial I(x_1)}\mathcal{L}(I,J) &= \int_{-\infty}^{\infty} J(x_1,x_2)\, p(x_1,x_2)\,[E(\mu|x_1,x_2)-\lambda]\, dx_2 - cp(x_1) \\ \frac{1}{p(x_1)}\frac{\partial}{\partial I(x_1)}\mathcal{L}(I,J) &= \int_{-\infty}^{\infty} J(x_1,x_2)\, p(x_2|x_1)\,[E(\mu|x_1,x_2)-\lambda]\, dx_2 - c \end{aligned}$$

By Theorem 3, $J(x_1,x_2) \geq 0 \iff [E(\mu|x_1,x_2)-\lambda] \geq 0$. Thus,

$$\frac{1}{p(x_1)}\frac{\partial}{\partial I(x_1)}\mathcal{L}(I,J) = \int_{-\infty}^{\infty} p(x_2|x_1)\max(E(\mu|x_1,x_2)-\lambda, 0)\, dx_2 - c$$

For clarity of the argument, define

$$\omega(x_1,x_2) =: \max(E(\mu|x_1,x_2)-\lambda,0)$$

and write

$$\frac{\partial}{\partial I(x_1)}\mathcal{L}(I,J) = \int_{-\infty}^{\infty} p(x_2|x_1)\,\omega(x_1,x_2)\, dx_2 - c$$

To prove the result, we will show that $\frac{\partial}{\partial I(x_1)}\mathcal{L}(I,J) \leq \frac{\partial}{\partial I(\hat{x}_1)}\mathcal{L}(I,J)$ if $x_1 \leq \hat{x}_1$.

Due to log-supermodularity, $E(\mu|x_1,x_2)$ is increasing with both x_1 and x_2, and therefore ω is increasing in x_1 and x_2. To complete the proof, it is enough to that, if f is log supermodular and $x_1 \leq \hat{x}_1$, then

$$\int p(x_2|x_1)\max(E(\mu|x_1,x_2)-\lambda,0)dx_2 \leq \int p(x_2|\hat{x}_1)\max(E(\mu|\hat{x}_1,x_2)-\lambda,0)dx_2$$

Let

$$
\begin{aligned}
u(x_2) &= \max(E(\mu|\hat{x}_1, x_2) - \lambda, 0) \\
v(x_2) &= \max(E(\mu|x_1, x_2) - \lambda, 0) \\
F(x_2) &= \int_{\infty}^{x_2} p(\tilde{x}_2|\hat{x}_1) d\tilde{x}_2 \\
G(x_2) &= \int_{\infty}^{x_2} p(\tilde{x}_2|x_1) d\tilde{x}_2
\end{aligned}
$$

In this notation, we want to show that

$$\int v\left(x_2\right) dG\left(x_2\right) \leq \int u\left(x_2\right) dF\left(x_2\right)$$

$E(\mu|x_1, x_2)$ is weakly increasing in both arguments by log supermodularity of $p(x_1, x_2, \mu)$, and therefore v and u are weakly increasing. Because $\hat{x}_1 > x_1$, $v \leq u$, and because F first-order dominates the distribution function G (see the Remark) the result follows from Lemma 1. ■

When there is a sufficient statistic, the optimal threshold policy can be stated with reference to the sufficient statistic in round two, just as for a single round of elimination.

Theorem 5 (Double elimination with memory: sufficient statistic) *Suppose the distribution p is log supermodular, and that σ is a sufficient statistic for μ. Given a cost c, let Δ_1, Δ_2 be the selection sets that solve (9). Then the selection sets can be written with threshold values $k_1, \bar{\sigma}$ such that*

$$x_1 \geq k_1 \textit{ for a.e. } x_1 \in \Delta_1$$

$$x_1 \leq k_1 \textit{ for a.e. } x_1 \notin \Delta_1$$

$$\textit{for a.e. } x_1 \in \Delta_1, \begin{cases} \sigma\left(x_1, x_2\right) \geq \bar{\sigma} \textit{ for a.e. } x_2 \in \Delta_2\left(x_1\right) \\ \\ \sigma\left(x_1, x_2\right) \leq \bar{\sigma} \textit{ for a.e. } x_2 \notin \Delta_2\left(x_1\right) \end{cases}$$

Proof: The characterization of Δ_1 is the same as in Theorem 4. We must show that the characterization of Δ_2 in Theorem 3 is equivalent to the one in this theorem. The posterior distribution of μ, given (x_1, x_2), is

$$\begin{aligned} p(\mu|x_1,x_2) &= \frac{p(x_1,x_2,\mu)}{\int p(x_1,x_2,\mu)\,d\mu} = \frac{g(x_1,x_2)\,\theta(\sigma(x_1,x_2),\mu)}{\int g(x_1,x_2)\,\theta(\sigma(x_1,x_2),\mu)\,d\mu} \\ &= \frac{\theta(\sigma(x_1,x_2),\mu)}{\int \theta(\sigma(x_1,x_2),\mu)\,d\mu} \end{aligned}$$

Therefore

$$\begin{aligned} E(\mu|x_1,x_2) &= \int \mu \left[\frac{\theta(\sigma(x_1,x_2),\mu)}{\int \theta(\sigma(x_1,x_2),\mu)\,d\mu}\right] d\mu \\ &= \bar{E}(\mu|\sigma(x_1,x_2)) \end{aligned}$$

Using Lemma 2, $\bar{E}(\mu|\sigma)$ is increasing in σ, and σ is increasing in (x_1, x_2). Choose $\bar{\sigma}$ so that $\bar{E}(\mu|\bar{\sigma}) = \lambda$. Then the characterization of Δ_2 above is equivalent to the characterization of Δ_2 in Theorem 3. ■

4.2 Double round with memory: comparisons

As shown in Theorem 3, the expected ability of the marginal survivor in an optimal selection scheme is λ at the end of round two. This is also the opportunity cost of reducing the number of survivors; it is the shadow price on the constraint that a fraction Λ of projects must survive. We now ask how the selection scheme and its efficacy change when the cost of keeping candidates in the pool increases. Given that mistakes are made at round one – some of the high-μ agents or projects are eliminated due to the randomness in x_1 – it is not entirely obvious how the ability of marginal survivors relates to the average ability in the group that survives.

The costliness of collecting information creates two burdens. First is the direct burden of paying the cost of round-one survivors in round two. Second, by eliminating some of the agents or projects too early, the selection process is less effective. We show that, if the cost of keeping survivors after round one increases, fewer will be kept, and the average ability of ultimate survivors, after round two, becomes smaller.

Theorem 6 (Double eliminations with memory: Higher cost leads to more stringent screening at round one, less stringent screening at round two, and lower average ability of survivors at the end.) *For each cost* c*, let* (Δ_1^c, Δ_2^c) *be the optimal selection sets for the double-elimination problem (9). Let* $\hat{c} > c$*. Then*

(1) If $0 < p\left(\Delta_1^{\hat{c}}\right) < 1$ *and* $0 < p\left(\Delta_1^c\right) < 1$*, then* $p\left(\Delta_1^{\hat{c}}\right) \leq p\left(\Delta_1^c\right)$*.*

(2) The expected ability of survivors in the selection scheme (Δ_1^c, Δ_2^c) *is larger (no smaller) than in the selection scheme* $\left(\Delta_1^{\hat{c}}, \Delta_2^{\hat{c}}\right)$*.*

(3) If p *is strictly log supermodular, the selection sets can be written with threshold values as in Theorem 4 where*

$$\begin{aligned} k_1(\hat{c}) &\geq k_1(c) \\ k_2(x_1, \hat{c}) &\leq k_2(x_1, c) \ \text{ at each } x_1 \in \Delta_1^c \cap \Delta_1^{\hat{c}} \end{aligned}$$

Proof: (1) Using Theorem 3, write the objective function as V, defined as

$$\begin{aligned} V(\Delta_1, \Delta_2; c) &= T(\Delta_1, \Delta_2) - cp(\Delta_1) \\ \text{where } T(\Delta_1, \Delta_2) &= \int \int_{\Delta_1} \int_{\Delta_2(x_1)} \mu p(x_1, x_2, \mu)\, dx_2 dx_1 d\mu \\ cp(\Delta_1) &= c \int \int_{\Delta_1} \int p(x_1, x_2, \mu)\, dx_2 dx_1 d\mu \end{aligned}$$

Because

$$\begin{aligned} V(\Delta_1^c, \Delta_2^c; c) &= T(\Delta_1^c, \Delta_2^c) - cp(\Delta_1^c) \geq T\left(\Delta_1^{\hat{c}}, \Delta_2^{\hat{c}}\right) - cp\left(\Delta_1^{\hat{c}}\right) \\ V\left(\Delta_1^{\hat{c}}, \Delta_2^{\hat{c}}; \hat{c}\right) &= T\left(\Delta_1^{\hat{c}}, \Delta_2^{\hat{c}}\right) - \hat{c}p\left(\Delta_1^{\hat{c}}\right) \geq T(\Delta_1^c, \Delta_2^c) - \hat{c}p(\Delta_1^c) \end{aligned}$$

it follows that

$$\begin{aligned} T(\Delta_1^c, \Delta_2^c) - T\left(\Delta_1^{\hat{c}}, \Delta_2^{\hat{c}}\right) &\geq c\left[p(\Delta_1^c) - p\left(\Delta_1^{\hat{c}}\right)\right] \\ T(\Delta_1^c, \Delta_2^c) - T\left(\Delta_1^{\hat{c}}, \Delta_2^{\hat{c}}\right) &\leq \hat{c}\left[p(\Delta_1^c) - p\left(\Delta_1^{\hat{c}}\right)\right] \end{aligned}$$

Subtracting,

$$0 \geq (c - \hat{c})\left[p(\Delta_1^c) - p\left(\Delta_1^{\hat{c}}\right)\right]$$

Thus $(c - \hat{c}) < 0 \implies p(\Delta_1^c) \geq p\left(\Delta_1^{\hat{c}}\right).$

(2) The total ability of survivors is $T(\Delta_1^c, \Delta_2^c)$ for each c. If (Δ_1^c, Δ_2^c) is optimal for c, and $\left(\Delta_1^{\hat{c}}, \Delta_2^{\hat{c}}\right)$ is optimal for $\hat{c}$, the two selection schemes yield the same number of survivors. Because $T(\Delta_1^c, \Delta_2^c) - cp(\Delta_1^c) \geq T\left(\Delta_1^{\hat{c}}, \Delta_2^{\hat{c}}\right) - cp\left(\Delta_1^{\hat{c}}\right)$ and $p(\Delta_1^c) - p\left(\Delta_1^{\hat{c}}\right) \geq 0$, it follows that

$$T(\Delta_1^c, \Delta_2^c) - T\left(\Delta_1^{\hat{c}}, \Delta_2^{\hat{c}}\right) \geq c\left[p(\Delta_1^c) - p\left(\Delta_1^{\hat{c}}\right)\right] \geq 0. \tag{18}$$

Dividing the left side by the probability of surviving both rounds, which is the same in both cases, the result follows.

(3) uses Theorem 4 and part (1) above.

$$\begin{aligned} p(\Delta_1^c) - p\left(\Delta_1^{\hat{c}}\right) &= \int \int_{k_1(c)} \int p(x_1, x_2, \mu)\, dx_2 dx_1 d\mu - \int \int_{k_1(\hat{c})} \int p(x_1, x_2, \mu)\, dx_2 dx_1 d\mu \\ &= \int \int_{k_1(c)}^{k_1(\hat{c})} \int p(x_1, x_2, \mu)\, dx_2 dx_1 d\mu \geq 0 \end{aligned}$$

The difference can only be positive if $k_1(\hat{c}) \geq k_1(c)$.

Then there exists at least one value x_1 such that $k_2(x_1, \hat{c}) \leq k_2(x_1, c)$, because otherwise the two schemes would not have the same numbers of survivors. Because

$$\hat{\lambda} \equiv E(\mu | x_1, k_2(x_1, \hat{c})) \leq E(\mu | x_1, k_2(x_1, c)) \equiv \lambda \tag{19}$$

at this x_1, it follows that $\hat{\lambda} \leq \lambda$, and therefore (19) holds at every $x_1 \in \Delta_1^c \cap \Delta_1^{\hat{c}}$, so that $k_2(x_1, \hat{c}) \leq k_2(x_1, c)$. ■

Because a single round of elimination is the extreme case where $c = 0$ and everyone is promoted or rejected after one round, we have the following corollary:

Corollary 3 (When cost is zero, a single round of elimination is optimal) *Conditional on the same number of survivors, the expected ability of survivors after the optimal single round of elimination is larger than the expected ability of survivors in any double round of elimination in which some agents or projects are eliminated at round one.*

5 Double Rounds of elimination without memory

When we say that the elimination scheme does not have memory, we mean that the selection set at round two cannot depend on x_1. Only the fact of survival is known from the first round. The selection problem can now be described as the choice of $\Delta_1 \subset \mathbf{R}$ and $\Delta_2 \subset \mathbf{R}$, where the selection at round two requires both $x_1 \in \Delta_1$ and $x_2 \in \Delta_2$.

The objective is still to maximize the expected μ among agents who survive both rounds, minus the cost that must be paid in the second round for survivors of the first round. We write the objective function as

$$\begin{aligned} V(\Delta_1, \Delta_2; c) &=: \int \int_{\Delta_1} \int_{\Delta_2} \mu p(x_1, x_2, \mu)\, dx_2 dx_1 d\mu \\ &\quad - c \int \int_{\Delta_1} \int \mu p(x_1, x_2, \mu)\, dx_2 dx_1 d\mu \end{aligned}$$

We write the number of survivors of both rounds as

$$S(\Delta_1, \Delta_2) =: \int \int_{\Delta_1} \int_{\Delta_2} p(x_1, x_2, \mu)\, dx_2 dx_1 d\mu$$

Then the objective is

$$\text{maximize } V(\Delta_1, \Delta_2; c) \text{ subject to } S(\Delta_1, \Delta_2) = \Lambda \tag{20}$$

Theorem 7 (Double elimination without memory: the optimal policy) *Let Δ_1, Δ_2 be selection sets that solve (20). Then there exists a number λ such that*

$$\begin{aligned} E(\mu | x_1, \Delta_2) - \frac{c}{p(\Delta_2 | x_1)} - \lambda &\geq 0 \textit{ for a.e. } x_1 \in \Delta_1 \\ E(\mu | x_1, \Delta_2) - \frac{c}{p(\Delta_2 | x_1)} - \lambda &\geq 0 \textit{ for a.e. } x_1 \notin \Delta_1 \end{aligned} \tag{21}$$

$$\begin{aligned} E(\mu | \Delta_1, x_2) - \lambda &\geq 0 \textit{ for a.e. } x_2 \in \Delta_2 \\ E(\mu | \Delta_1, x_2) - \lambda &\leq 0 \textit{ for a.e. } x_2 \notin \Delta_2 \end{aligned} \tag{22}$$

Proof: In order to characterize the solution, it is again convenient to reformulate the objective as a Lagrange function where the choice variables are indicator values $I(x_1) \in \{0,1\}$ for each $x_1 \in \mathbf{R}$ and $J(x_2) \in \{0,1\}$ for each $(x_1,x_2) \in \mathbf{R}\times\mathbf{R}$ such that $I(x_1) = 1$. There is again a one-to-one relationship between the indicator functions I, J and the selection sets Δ_1, Δ_2, now given by

$$\Delta_1 = \{x_1 \in \mathbf{R} : I(x_1) = 1\}$$

$$\Delta_2 = \{x_2 \in \mathbf{R} : J(x_2) = 1\}$$

The Lagrange function to be maximized is

$$\begin{aligned}
&\mathcal{L}(I,J) = \int\int I(x_1)\int J(x_2) \quad \mu\, p(x_1,x_2,\mu)\, dx_2 dx_1 d\mu \\
&-c\int I(x_1)\int\int p(x_1,x_2,\mu)\, dx_2 dx_1 d\mu \\
&-\lambda\left[\int\int I(x_1)\int J(x_2)\, p(x_1,x_2,\mu)\ dx_2 dx_1 d\mu - \Lambda\right]
\end{aligned}$$

The first-order conditions are $S(\Delta_1,\Delta_2) = \Lambda$ and for each (x_1,x_2),

$$\begin{cases} \dfrac{\partial}{\partial I(x_1)}\mathcal{L}(I,J) \geq 0 & \text{if} \quad I(x_1) = 1 \\[2ex] \dfrac{\partial}{\partial I(x_1)}\mathcal{L}(I,J) \leq 0 & \text{if} \quad I(x_1) = 0 \end{cases} \tag{23}$$

$$\begin{cases} \dfrac{\partial}{\partial J(x_2)}\mathcal{L}(I,J) \geq 0 & \text{if} \quad J(x_2) = 1 \\[2ex] \dfrac{\partial}{\partial J(x_2)}\mathcal{L}(I,J) \leq 0 & \text{if} \quad J(x_2) = 0 \end{cases} \tag{24}$$

$$\begin{aligned}
\frac{\partial}{\partial I(x_1)}\mathcal{L}(I,J) = \int\int J(x_2)\,\mu\int_{-\infty}^{\infty} p(x_1,x_2,\mu)\, dx_2 d\mu - c\int\int_{-\infty}^{\infty} p(x_1,x_2,\mu)\, dx_2 d\mu \\
- \lambda\int\int J(x_2)\, p(x_1,x_2,\mu)\, dx_2 d\mu
\end{aligned}$$

For each $\Delta_1 \subset \mathbf{R}, \Delta_2 \subset \mathbf{R}$, we use the notation

$$\begin{aligned}
E(\mu|x_1,x_2) &= \int \mu \frac{p(x_1,x_2,\mu)}{\int p(x_1,x_2,\mu)\,d\mu} d\mu \\
&= \int \mu p(\mu|x_1,x_2)\,d\mu \\
E(\mu|x_1,\Delta_2) &= \int_{\Delta_2}\int \mu \frac{p(x_1,x_2,\mu)}{\int_{\Delta_2}\int p(x_1,x_2,\mu)\,d\mu dx_2} d\mu dx_2 \\
&= \int_{\Delta_2} E(\mu|x_1,x_2)\frac{p(x_2|x_1)}{p(\Delta_2|x_1)} dx_2 \\
E(\mu|\Delta_1,x_2) &= \int_{\Delta_1}\int \frac{p(x_1,x_2,\mu)}{\int_{\Delta_1}\int p(x_1,x_2,\mu)\,d\mu dx_1} d\mu dx_1 \\
&= \int_{\Delta_1} E(\mu|x_1,x_2)\frac{p(x_1|x_2)}{p(\Delta_1|x_2)} dx_1
\end{aligned}$$

$$\begin{aligned}
\frac{\partial}{\partial I(x_1)}\mathcal{L}(I,J) &= \int J(x_2)\,p(x_1,x_2)\int \mu \frac{p(x_1,x_2,\mu)}{p(x_1,x_2)} d\mu dx_2 \\
&\quad -cp(x_1) - \lambda \int J(x_2)\,p(x_1,x_2)\,dx_2 \\
&= \int J(x_2)\,p(x_1,x_2)\,E(\mu|x_1,x_2)\,dx_2 - cp(x_1) - \lambda p(x_1,\Delta_2)
\end{aligned}$$

$$\frac{\partial}{\partial I(x_1)}\mathcal{L}(I,J)\times\frac{1}{p(x_1)} = \int J(x_2)\,p(x_2|x_1)\,E(\mu|x_1,x_2)\,dx_2 - c - \lambda\int J(x_2)\,p(x_2|x_1)\,dx_2 \tag{25}$$

$$\begin{aligned}
&\frac{\partial}{\partial I(x_1)}\mathcal{L}(I,J)\times\frac{1}{p(x_1,\Delta_2)} \\
&= \int J(x_2)\frac{p(x_1,x_2)}{p(x_1,\Delta_2)}E(\mu|x_1,x_2)\,dx_2 - c\frac{p(x_1)}{p(x_1,\Delta_2)} - \lambda \\
&= \int_{\Delta_2}\frac{p(x_2|x_1)}{p(\Delta_2|x_1)}E(\mu|x_1,x_2)\,dx_2 - c\frac{1}{p(\Delta_2|x_1)} - \lambda \\
&= E(\mu|x_1,\Delta_2) - \frac{c}{p(\Delta_2|x_1)} - \lambda
\end{aligned} \tag{26}$$

$$\begin{aligned}
&\mathcal{L}(I,J) = \int\int I(x_1)\int J(x_2) \quad \mu \quad p(x_1,x_2,\mu)\,dx_2dx_1d\mu \\
&-c\int I(x_1)\int \quad f(x_1,\mu)\,h(\mu)\,dx_1d\mu \\
&-\lambda\int\int I(x_1)\int J(x_2) \quad p(x_1,x_2,\mu)\,dx_2dx_1d\mu
\end{aligned}$$

$$
\begin{aligned}
\frac{\partial}{\partial J(x_2)}\mathcal{L}(I,J) &= \int I(x_1)\int \mu \;\; p(x_1,x_2,\mu)\,d\mu dx_1 - \lambda \int I(x_1)\int p(x_1,x_2,\mu)\,d\mu dx_1 \\
&= p(\Delta_1,x_2)\{E(\mu|\Delta_1,x_2)-\lambda\} \qquad (27)
\end{aligned}
$$

The theorem then follows from (23) and (24), and (26) and (27). ■

As with memory, if the probability distributions satisfy a monotonicity property, then the optimal selection sets can be expressed using threshold values, and the threshold values depend on the cost.

Theorem 8 (Double elimination without memory: promotion using threshold values) *Suppose the distribution p is log supermodular. Given a cost c, let Δ_1, Δ_2 be the selection sets that solve (20). Then the selection sets can be written with threshold values $k_1(c), k_2(c)$ such that*

$$
\begin{aligned}
&x_1 \geq k_1(c) \text{ for a.e. } x_1 \in \Delta_1 \\
&x_1 \leq k_1(c) \text{ for a.e. } x_1 \notin \Delta_1 \\
&x_2 \geq k_2(c) \text{ for a.e. } x_2 \in \Delta_2 \\
&x_2 \leq k_2(c) \text{ for a.e. } x_2 \notin \Delta_2
\end{aligned}
$$

This is proved as in Theorem 4.

Finally, we can make a qualitative statement about the stringency of screening at round one. Even without log supermodularity, we can conclude that higher cost should lead to fewer survivors of round one, and that the more stringent policy reduces the average ability of ultimate survivors. With log supermodularity, the more stringent policy takes the form of a higher threshold value for survival at round one, and a corresponding lower threshold at round two, in order to ensure that there are enough ultimate survivors.

Theorem 9 [Double eliminations without memory: Higher cost leads to more stringent screening at round one, less stringent screening at round two, and survivors of lower average ability at the end.]

For each cost c, let Δ_1^c, Δ_2^c be the selection sets that solve (20). Let $\hat{c} > c$. Then

(1) If $0 < p\left(\Delta_1^{\hat{c}}\right), p\left(\Delta_1^c\right) < 1$, then $p\left(\Delta_1^c\right) \geq p\left(\Delta_1^{\hat{c}}\right)$.

(2) The expected ability of survivors in the selection scheme $\left(\Delta_1^c, \Delta_2^c\right)$ is larger than in the selection scheme $\left(\Delta_1^{\hat{c}}, \Delta_2^{\hat{c}}\right)$.

(3) If p is strictly log supermodular, the selection sets can be written with threshold values as in Theorem 8, where

$$k_1(\hat{c}) > k_2(c) \text{ and } k_2(\hat{c}) \leq k_2(c).$$

The proof of this theorem is the same as the proof of Theorem 6 except that we must use the definition of V in the problem (20) where the selection criterion in round two does not depend on x_1.

Finally, we ask whether selection standards should become tougher or more lenient over time. That is, should the standard in the second round be lower or higher than in the first round? We did not address this question for the selection scheme with memory, because the selection standards in the second round depend on the signal from the first round.

It is convenient to return to the underlying densities f and h, instead of the density function p. Let ϕ be the hazard rate of f:

$$\phi(k, \mu) \equiv \frac{f(k, \mu)}{[1 - F(k, \mu)]}$$

The assumption under which we can rank the thresholds in the two rounds is that the cross partial is positive:

$$\frac{\partial^2}{\partial \mu \partial k} \log \phi(k, \mu) > 0 \tag{28}$$

Lemma 3 *Let $k_1, k_2 \in \mathbf{R}$. Suppose that (28) holds. Then*

$$E(\mu | k_1, (k_2, \infty)) \left\{ \begin{array}{c} > \\ = \\ < \end{array} \right\} E(\mu | (k_1, \infty), k_2) \quad \begin{array}{l} \text{if } k_1 > k_2 \\ \text{if } k_1 = k_2 \\ \text{if } k_1 < k_2 \end{array}$$

Proof: Write

$$
\begin{aligned}
p\left(\left(k_1,\infty\right),k_2\right) &= \int_{k_1}\int f\left(x_1,\mu\right) f\left(k_2,\mu\right) h\left(\mu\right) d\mu dx_1 \\
&= \int \phi\left(k_2,\mu\right)\left[1-F\left(k_1,\mu\right)\right]\left[1-F\left(k_2,\mu\right)\right] h\left(\mu\right) d\mu \\
p\left(k_1,\left(k_2,\infty\right)\right) &= \int \phi\left(k_1,\mu\right)\left[1-F\left(k_1,\mu\right)\right]\left[1-F\left(k_2,\mu\right)\right] h\left(\mu\right) d\mu
\end{aligned}
$$

$$
\begin{aligned}
E\left(\mu|\left(k_1,\infty\right),k_2\right) &= \int \mu \frac{\phi\left(k_2,\mu\right)\left[1-F\left(k_1,\mu\right)\right]\left[1-F\left(k_2,\mu\right)\right] h\left(\mu\right)}{p\left(\left(k_1,\infty\right),k_2\right)} d\mu \\
&= \int \mu g^{k_2}\left(\mu\right) d\mu
\end{aligned}
$$

where g^{k_2} is a probability distribution defined for each μ by

$$
g^{k_2}\left(\mu\right) = \frac{\phi\left(k_2,\mu\right)\left[1-F\left(k_1,\mu\right)\right]\left[1-F\left(k_2,\mu\right)\right] h\left(\mu\right)}{p\left(\left(k_1,\infty\right),k_2\right)}
$$

Similarly,

$$
\begin{aligned}
E\left(\mu|k_1,\left(k_2,\infty\right)\right) &= \int \mu \frac{\phi\left(k_1,\mu\right)\left[1-F\left(k_1,\mu\right)\right]\left[1-F\left(k_2,\mu\right)\right] h\left(\mu\right)}{p\left(k_1,\left(k_2,\infty\right)\right)} d\mu \\
&= \int \mu g^{k_1}\left(\mu\right) d\mu
\end{aligned}
$$

where g^{k_1} is a probability distribution defined for each μ by

$$
g^{k_1}\left(\mu\right) = \frac{\phi\left(k_1,\mu\right)\left[1-F\left(k_1,\mu\right)\right]\left[1-F\left(k_2,\mu\right)\right] h\left(\mu\right)}{p\left(k_1,\left(k_2,\infty\right)\right)}
$$

If $k_1 = k_2$, then $g^{k_1}\left(\mu\right) = g^{k_2}\left(\mu\right)$, hence $E\left(\mu|k_1,\left(k_2,\infty\right)\right) = E\left(\mu|\left(k_1,\infty\right),k_2\right)$.

If the cross partial (28) is positive, the following shows that the ratio $g^{k_1}\left(\mu\right)/g^{k_2}\left(\mu\right)$ is increasing if $k_1 > k_2$, and therefore g^{k_1} stochastically dominates g^{k_2}. The reverse holds if $k_2 > k_1$.

$$
\begin{aligned}
\frac{\partial}{\partial\mu}\log\frac{g^{k_1}\left(\mu\right)}{g^{k_2}\left(\mu\right)} &= \frac{\partial}{\partial\mu}\log\frac{\phi\left(k_1,\mu\right)}{\phi\left(k_2,\mu\right)} = \frac{\partial}{\partial\mu}\log\phi\left(k_1,\mu\right) - \frac{\partial}{\partial\mu}\log\phi\left(k_2,\mu\right) \\
&= \int_{k_2}^{k_1}\frac{\partial^2}{\partial\mu\partial k}\log\phi\left(k,\mu\right) dk \qquad \blacksquare
\end{aligned}
$$

This lemma allows us to state the following theorem:

Theorem 10 (Without memory, screening should become less stringent) *For a given cost c, suppose the selection sets that solve (20) can be written with threshold values $(k_1(c), k_2(c))$ as in Theorem 8. Suppose that the cross partial of the derivative of the logarithm of ϕ is positive. Then for each c, $k_1(c) > k_2(c)$.*

Proof: Theorem 7 implies that

$$E\left(\mu | k_1(c), (k_2(c), \infty)\right) - \frac{c}{p\left((k_2(c), \infty) | k_1(c)\right)} - E\left(\mu | (k_1(c), \infty), k_2(c)\right) = 0$$

If $c > 0$, the result follows from Lemma 3 because $E\left(\mu | k_1(c), (k_2(c), \infty)\right) > E\left(\mu | (k_1(c), \infty), k_2(c)\right)$.

■

References

[1] Athey, S. 2002. Monotone Comparative Statics Under Uncertainty. *Quarterly Journal of Economics* 117: 187-223.

[2] Milgrom, P.R. 1981. Good News and Bad News: Representation Theorems and Applications, *The Bell Journal of Economics.* 12:380-91

[3] Milgrom, P. and C. Shannon. 1994. Monotone Comparative Statics, *Econometrica* 62(1):157-80.

[4] Milgrom, P.R. and R. J. Weber, R.J. 1982. A theory of auctions and competitive bidding,*Econometrica* 50: 1089-1122.

[5] Scotchmer, S. 2008. Risk Taking and Gender in Hierarchies. *Theoretical Economics* 3:499-524.

[6] Shaked, M. and Shanthikumar, J.G. 2006. *Stochastic Orders.* Springer (Springer Series in Statistics): New York.

[7] Topkis,D.M. 1978. Minimizing a Submodular Function on a Lattice, *Mathematics of Operations Research*. 305-321.

Appendix: Theorem 1, Proof of Existence

To show: for each Λ, there exists a set S with the property that:

1. $\nu(S) = \Lambda$,

2. There exists a finite number α such that $g \geq \alpha$ on S, and $g \leq \alpha$ on the complement.

To this end, for each $a \in (-\infty, \infty)$, define $t(a) = \nu\{g \leq a\}$, and define right and left limits:

$t_{-}(a) := \lim_{n\to\infty} t(a - \frac{1}{n}) = \lim_{n\to\infty} \nu\{g \leq a - \frac{1}{n}\} = \nu(\{g < a\})$

$t_{+}(a) := \lim_{n\to\infty} t(a + \frac{1}{n}) = \lim_{n\to\infty} \nu\{g \leq a + \frac{1}{n}\} = \nu(\{g \leq a\}) = t(a).$

Because t is monotonic increasing, both the left limit $t_{-}(a)$ and right limit $t_{+}(a)$ exist. Moreover, t is right continuous, but may have jumps. Let $t_{+}(a) - t_{-}(a) = \nu(\{g = a\}) \geq 0$ be the jump at a.

Let $\lambda = 1 - \Lambda$, hence $\lambda \in (0, 1)$. Define $\alpha = \inf\{a \mid t(a) \geq \lambda\}$. Then α is finite and $t_{-}(\alpha) \leq \lambda \leq t_{+}(\alpha) = t(\alpha)$. The jump at α is $t(\alpha) - t_{-}(\alpha)$. There are two cases:

1. If there is zero jump at α, choose $S = \{g > \alpha\}$. Then $\nu(S) = 1 - t(\alpha) = 1 - \lambda = \Lambda$. Clearly $g \leq \alpha$ on the complement of S.

2. If there is a positive jump at α, $\nu(\{g = \alpha\} = t(\alpha) - t_{-}(\alpha) > 0$. Then $0 \leq \lambda - t_{-}(\alpha) \leq t(\alpha) - t_{-}(\alpha)$. Since ν is atomless, there exists a subset $P_1 \subset \{g = \alpha\}$ with measure $\nu(P_1) = \lambda - t_{-}(\alpha)$. Let $P_2 = \{g = \alpha\} \backslash P_1$, $P^{+} = \{g > \alpha\}$, $P^{-} = \{g < \alpha\}$, then (P^{+}, P^{-}, P_1, P_2) is a partitioning of the whole space. Choose $S = P^{+} \cup P_2$. Then the complement of S is $P^{-} \cup P_1$, so $g \leq \alpha$ on the complement, and $\nu(\{g < \alpha\}) + \nu(P_1) = t_{-}(\alpha) + \lambda - t_{-}(\alpha) = \lambda$. Therefore and $\nu(S) = 1 - \lambda = \Lambda$ and $g \geq \alpha$ on S. ■