

COMPARATIVE STATICS, INFORMATIVENESS, AND
THE INTERVAL DOMINANCE ORDER

By John K.-H. Quah and Bruno Strulovici

Abstract: We identify a natural way of ordering functions, which we call the *interval dominance order*, and show that this concept is useful in the theory of monotone comparative statics and also in statistical decision theory. This ordering on functions is weaker than the standard one based on the single crossing property (Milgrom and Shannon, 1994) and so our monotone comparative statics results apply in some settings where the single crossing property does not hold. For example, they are useful when examining the comparative statics of optimal stopping time problems. We also show that certain basic results in statistical decision theory which are important in economics - specifically, the complete class theorem of Karlin and Rubin (1956) and the results connected with Lehmann's (1988) concept of informativeness - generalize to payoff functions that obey the interval dominance order.

Keywords: single crossing property, interval dominance order, supermodularity, comparative statics, optimal stopping time, complete class theorem, statistical decision theory, informativeness.

JEL Classification Numbers: C61, D11, D21, F11, G11.

Authors' Emails: john.quah@economics.ox.ac.uk bruno.strulovici@economics.ox.ac.uk

Acknowledgments: We would like to thank Ian Jewitt for many stimulating conversations. We also received many helpful comments from seminar and conference participants in Budapest, Exeter, Kos, Michigan, Northwestern, Oxford, Singapore, and Warwick. In particular, we would like to thank Rabah Amir, Alan Beggs, Eddie Dekel, Juan-Jose Ganuza, Paul Milgrom, Leonard Mirman, Herakles Polemarchakis, Edward Schlee, and Aleksey Tetenov.

1. INTRODUCTION

The single crossing property is of fundamental importance in the theory of monotone comparative statics because it is a central concept of the theory and is widely used in applications.¹ Let X be a subset of the real line R , and $\{f(\cdot, s)\}_{s \in S}$ a family of functions mapping X to R and parameterized by s in $S \subseteq R$. We say that this family is ordered by the *single crossing property* if for all $s'' > s'$ and $x'' > x'$, the following holds:

$$f(x'', s') - f(x', s') \geq (>) 0 \implies f(x'', s'') - f(x', s'') \geq (>) 0. \quad (1)$$

To keep our discussion simple, assume that $f(\cdot, s)$ has a unique maximum in X for all s . The significance of the single crossing property (SCP) arises from this result (Milgrom and Shannon (1994)): *if the family $\{f(\cdot, s)\}_{s \in S}$ is ordered by the single crossing property, then $\operatorname{argmax}_{x \in X} f(x, s)$ is increasing in s .*

It is trivial to see that SCP is also necessary for comparative statics in the following sense: if we require $\operatorname{argmax}_{x \in Y} f(x, s'') \geq \operatorname{argmax}_{x \in Y} f(x, s')$ for all $Y \subset X$ whenever $s'' > s'$, then (1) must hold. (This is true since we could let Y be any set consisting of two points in X .) In other words, SCP is necessary if we require the monotone comparative statics conclusion to be robust to any change in the domain of the objective functions.

However, SCP is *not* necessary for monotone comparative statics if we only require $\operatorname{argmax}_{x \in Y} f(x, s'') \geq \operatorname{argmax}_{x \in Y} f(x, s')$ for $Y = X$ or for Y belonging to a particular subcollection of the subsets of X . To see this, consider Figures 1 and 2. In both cases, we have $\operatorname{argmax}_{x \in X} f(x, s'') \geq \operatorname{argmax}_{x \in X} f(x, s')$; furthermore, $\operatorname{argmax}_{x \in Y} f(x, s'') \geq \operatorname{argmax}_{x \in Y} f(x, s')$ where Y is any closed interval contained in X . In Figure 1, SCP

¹Early contributions to the literature on monotone comparative statics include Topkis (1978), Milgrom and Roberts (1990), Vives (1990), and Milgrom and Shannon (1994). A textbook treatment can be found in Topkis (1998). Ashworth and Bueno de Mesquita (2006) discusses applications in political science.

is satisfied (specifically, (1) is satisfied) but this is not true in Figure 2. Consider the points x'' and x' as depicted in Figure 2. We see that $f(x', s') = f(x'', s')$ but $f(x'', s'') < f(x', s'')$, violating SCP.

The first objective of this paper is to develop a new way of ordering functions that guarantees monotone comparative statics, whether the situation is like the one depicted in Figure 1 or the one depicted in Figure 2. We call this new order the *interval dominance order*. This order is more general than the one based on the single crossing property. We show that it holds in some significant situations where SCP does not hold; at the same time it retains many of the nice comparative statics properties associated with SCP.

The second objective of this paper is to bridge the gap between the literature on monotone comparative statics and the closely related literature in statistical decision theory on informativeness. We refer in particular to Lehmann's concept of informativeness (1988) which in turn builds on the complete class theorem of Karlin and Rubin (1956).² In that setting, the state of the world is unknown and the agent has to take an action based on an observed signal, which conveys information on the true state. Karlin and Rubin identify conditions under which optimal decision rules (in some well-defined sense) must be monotone, i.e., the agent's action is higher when he receives a higher signal.³ Lehmann shows how we may compare the informativeness of two families of signals (or *experiments*) when the optimal decision rules are monotone.

A crucial assumption in the Karlin-Rubin and Lehmann theorems is that the agent's payoff in state s when she takes action x , which we write as $f(x, s)$, has the following property: $f(\cdot, s)$ is a quasiconcave function of x , achieving a maximum at $\bar{x}(s)$, with \bar{x} increasing in s (like the situation depicted in Figure 2). However,

²Economic applications of Lehmann's concept of informativeness can be found in Persico (2000), Athey and Levin (2001), Levin (2001), Bergemann and Valimaki (2002), and Jewitt (2006). For a recent application of Karlin and Rubin's complete class theorem, see Manski (2005).

³In other words, monotone decision rules form a *complete class*.

their results do not cover the case where $\{f(\cdot, s)\}_{s \in S}$ is ordered by the single crossing property. Amongst other things, SCP allows for non-quasiconcave payoff functions (see Figure 2); indeed this feature is crucial to many of its economics applications. In short, the standard results on comparative informativeness accommodates the situation depicted in Figure 2 but not that in Figure 1, whereas the standard results on comparative statics accommodate the situation in Figure 1 but not that in Figure 2.⁴

We generalize the Karlin-Rubin and Lehmann results by showing that their conclusions hold even when the payoff functions $\{f(\cdot, s)\}_{s \in S}$ satisfy the interval dominance order. In this way, we obtain a single condition on the payoff functions that is useful for *both* comparative statics and comparative informativeness, so results in one category extend seamlessly into results in the other.

The rest of this paper is organized as follows. In Section 2, we define the interval dominance order, explore its properties, and develop a comparative statics theorem. Section 3 is devoted to applications. In Section 4 we show that the concept of IDO can be easily extended to settings with uncertainty and that it is useful in that context. Lehmann's concept of informativeness is introduced in Section 5 and we demonstrate its relevance for payoff functions obeying the interval dominance order. Finally, we generalize Karlin and Rubin's complete class theorem in Section 6.

2. THE INTERVAL DOMINANCE ORDER

We begin by showing how a situation like that depicted in Figure 2, involving a violation of SCP, can arise naturally in an economic setting.

Example 1. Consider a firm producing some good whose price we assume is fixed at 1 (either because of market conditions or for some regulatory reason). It has

⁴Our observation that there is a distinction between the two types of payoff conditions was first highlighted in Jewitt (2006), which also discusses the significance of this distinction in applications.

to decide on the production capacity (x) of its plant. Assume that a plant with production capacity x costs Dx , where $D > 0$. Let s be the state of the world, which we identify with the demand for the good. The marginal cost of producing the good in state s is $c(s)$. We assume that, for all s , $D + c(s) < 1$. The firm makes its capacity decision *before* the state of the world is realized and its production decision *after* the state is revealed.

Suppose it chooses capacity x and the realized state of the world (and thus realized demand) is $s \geq x$. In this case, the firm should produce up to its capacity, so that its profit $\Pi(x, s) = x - c(s)x - Dx$. On the other hand, if $s < x$, the firm will produce (and sell) s units of the good, giving it a profit of $\Pi(x, s) = s - c(s)s - Dx$. It is easy to see that $\Pi(\cdot, s)$ is increasing linearly for $x \leq s$ and thereafter declines, linearly with a slope of $-D$. Its maximum is achieved at $x = s$, with $\Pi(s, s) = (1 - c(s) - D)s$. Suppose $s'' > s'$ and $c(s'') > c(s')$; in other words, the state with higher demand also has higher marginal cost. Then it is clear that the situation depicted in Figure 2 can arise, with $f(\cdot, s') = \Pi(\cdot, s')$ and $f(\cdot, s'') = \Pi(\cdot, s'')$.

Let X be a subset of R and f and g two real-valued functions defined on X . We say that g dominates f by the *single crossing property* (which we denote by $g \succeq_{SC} f$) if for all x'' and x' such that $x'' > x'$, the following holds:

$$f(x'') - f(x') \geq (>) 0 \implies g(x'') - g(x') \geq (>) 0. \quad (2)$$

A family of real-valued functions $\{f(\cdot, s)\}_{s \in S}$, defined on X and parameterized by s in $S \subset R$ is referred to as an *SCP family* if the functions are ordered by SCP, i.e., whenever $s'' > s'$, we have $f(\cdot, s'') \succeq_{SC} f(\cdot, s')$.

In Figure 2, we have $\operatorname{argmax}_{x \in R_+} f(x, s'') > \operatorname{argmax}_{x \in R_+} f(x, s')$ even though $f(\cdot, s'')$ does not dominate $f(\cdot, s')$ by SCP. Notice, however, that violations of (2) can only occur if we compare points x' and x'' on opposite sides of the maximum point of $f(\cdot, s')$. This suggests that a possible way of weakening SCP, while retaining

comparative statics, is to require (2) to hold only for a certain collection of pairs $\{x', x''\}$, rather than all possible pairs.

The set J is an *interval of X* if, whenever x' and x'' are in J , any element x in X such that $x' \leq x \leq x''$ is also in J .⁵ Let f and g be two real-valued functions defined on X . We say that g dominates f by the *interval dominance order* (or, for short, g I-dominates f , with the notation $g \succeq_I f$) if (2) holds for x'' and x' such that $x'' > x'$ and $f(x'') \geq f(x)$ for all x in the interval $[x', x''] = \{x \in X : x' \leq x \leq x''\}$.

Clearly, the interval dominance order (IDO) is weaker than ordering by SCP. For example, in Figure 2, $f(\cdot, s'')$ I-dominates $f(\cdot, s')$ but $f(\cdot, s'')$ does not dominate $f(\cdot, s')$ by SCP.

For many results in the paper, we shall impose a mild regularity condition on the objective function. A function $f : X \rightarrow R$ is said to be *regular* if $\operatorname{argmax}_{x \in [x', x'']} f(x)$ is nonempty for any points x' and x'' with $x'' > x'$. Suppose the set X is such that $X \cup [x', x'']$ is always closed, and thus compact, in R (with the respect to the Euclidean topology). This is true, for example, if X is finite, if it is closed, or if it is a (not necessarily closed) interval. Then f is regular if it is upper semi-continuous with respect to the relative topology on X .

We are now ready to examine the relationship between the interval dominance order and monotone comparative statics. Theorem 1 gives the precise sense in which IDO is both sufficient and necessary for monotone comparative statics. To deal with the possibility of multiple maxima we need a way of ordering sets. The standard way of ordering sets in this context is the *strong set order* (see Topkis (1998)). Let S' and S'' be two subsets of R . We say that S'' dominates S' in the strong set order, and write $S'' \geq S'$ if for any x'' in S'' and x' in S' , we have $\max\{x'', x'\}$ in S'' and

⁵Note that X need not be an interval in the conventional sense, i.e., X need not be, using our terminology, an *interval of R* . Furthermore, the fact that J is an interval of X does not imply that it is an interval of R . For example, if $X = \{1, 2, 3\}$, then $J = \{1, 2\}$ is an interval of X , but of course neither X nor J are intervals of R .

$\min\{x'', x'\}$ in S' . Suppose that S'' and S' both contain their largest and smallest elements. Then it is clear that if $S'' \geq S'$, the largest (smallest) element in S'' is greater than the largest (smallest) element in S' .⁶

THEOREM 1: *Suppose that f and g are real-valued functions defined on $X \subset \mathbb{R}$ and $g \succeq_I f$. Then the following property holds:*

$$(\star) \quad \operatorname{argmax}_{x \in J} g(x) \geq \operatorname{argmax}_{x \in J} f(x) \text{ for any interval } J \text{ of } X.$$

Furthermore, if property (\star) holds and g is regular, then $g \succeq_I f$.

Proof: Assume that g I-dominates f and that x'' is in $\operatorname{argmax}_{x \in J} f(x)$ and x' is in $\operatorname{argmax}_{x \in J} g(x)$. We need only consider the case where $x'' > x'$. Since x'' is in $\operatorname{argmax}_{x \in J} f(x)$, we have $f(x'') \geq f(x)$ for all x in $[x', x''] \subseteq J$. Since $g \succeq_I f$, we also have $g(x'') \geq g(x')$; thus x'' is in $\operatorname{argmax}_{x \in J} g(x)$. Furthermore, $f(x'') = f(x')$ so that x' is in $\operatorname{argmax}_{x \in J} f(x)$. If not, $f(x'') > f(x')$ which implies (by the fact that $g \succeq_I f$) that $g(x'') > g(x')$, contradicting the assumption that g is maximized at x' .

To prove the other direction, we assume that there is an interval $[x', x'']$ such that $f(x'') \geq f(x)$ for all x in $[x', x'']$. This means that x'' is in $\operatorname{argmax}_{x \in [x', x'']} f(x)$. There are two possible violations of IDO. One possibility is that $g(x') > g(x'')$; in this case, by the regularity of g , the set $\operatorname{argmax}_{x \in [x', x'']} g(x)$ is nonempty but does not contain x'' , which violates (\star) . Another possible violation of IDO occurs if $g(x'') = g(x')$ but $f(x'') > f(x')$. In this case, the set $\operatorname{argmax}_{x \in [x', x'']} g(x)$ either contains x' , which violates (\star) since $\operatorname{argmax}_{x \in [x', x'']} f(x)$ does not contain x' , or it does not contain x'' , which also violates (\star) . QED

For the interval dominance order to be useful in applications, it helps to have a simple way of checking that the property holds. For this purpose, the next result is crucial.

⁶Throughout this paper, when we say that something is ‘greater’ or ‘increasing’, we mean to say that it is greater or increasing in the *weak* sense. Most of the comparisons in this paper are weak, so this convention makes sense. When we are making a strict comparison, we shall say so explicitly, as in ‘strictly higher’, ‘strictly increasing’, etc.

PROPOSITION 1: *Suppose that X is an interval of R and the functions $f, g : X \rightarrow R$ are absolutely continuous on compact intervals in X (and thus f and g are differentiable a.e.). If there is an increasing and positive function $\alpha : X \rightarrow R$ such that $g'(x) \geq \alpha(x)f'(x)$ a.e., then $g \succeq_I f$.*

If the function α in Proposition 1 is a constant $\bar{\alpha}$, then we obtain $g(x'') - g(x') \geq \bar{\alpha}(f(x'') - f(x'))$, which implies $g \succeq_{sc} f$. When α is not a constant, the functions f and g in Proposition 2 need not be related by SCP, as the following example shows.

Let $f : [0, M] \rightarrow R$ be a differentiable and quasiconcave function, with $f(0) = 0$ and a unique maximum at x^* in $(0, M)$. Let $\alpha : [0, M] \rightarrow R$ be given by $\alpha(x) = 1$ for $x \leq x^*$ and $\alpha(x) = 1 + (x - x^*)$ for $x > x^*$. Consider $g : [0, M]$ satisfying $g(0) = f(0) = 0$ with $g'(x) = \alpha(x)f'(x)$ (as in Proposition 1). Then it is clear that $g(x) = f(x)$ for $x \leq x^*$ and $g(x) < f(x)$ for $x > x^*$. The function g is also quasiconcave with a unique maximum at x^* and g I-dominates f , but g does not dominate f by SCP.

Proposition 1 is a consequence of the following lemma.

LEMMA 1: *Suppose $[x', x'']$ is a compact interval of R and α and h are real-valued functions defined on $[x', x'']$, with h integrable and α increasing (and thus integrable as well). If $\int_x^{x''} h(t)dt \geq 0$ for all x in $[x', x'']$, then*

$$\int_{x'}^{x''} \alpha(t)h(t)dt \geq \alpha(x') \int_{x'}^{x''} h(t)dt. \quad (3)$$

Proof: We confine ourselves to the case where α is an increasing and differentiable function. If we can establish (3) for such functions, then we can extend it to all increasing functions α since any such function can be approximated by an increasing and differentiable function.

The function $H(t) = \alpha(t) \int_t^{x''} h(z)dz$ is absolutely continuous and thus differentiable a.e.; by the fundamental theorem of calculus, we have $H(x'') - H(x') =$

$\int_{x''}^{x'} H'(t)dt$. Note that $H(x'') = 0$ and that by the product rule,

$$H'(t) = \alpha'(x) \int_t^{x''} h(z)dz - \alpha(t)h(t).$$

So

$$-H(x') = -\alpha(x') \int_{x'}^{x''} h(t)dt = \int_{x'}^{x''} \alpha'(x) \left(\int_t^{x''} h(z)dz \right) dt - \int_{x'}^{x''} \alpha(t)h(t)dt.$$

Note that the first term on the right of this equation is always nonnegative by assumption and so we obtain (3). QED

Proof of Proposition 1: Consider x'' and x' in X such that $x'' > x'$ and assume that $f(x) \leq f(x'')$ for all x in $[x', x'']$. Since f is absolutely continuous on $[x', x'']$, $f(x'') - f(x) = \int_x^{x''} f'(t)dt$ (with an analogous expression for g). We then have

$$\int_x^{x''} g'(t)dt \geq \int_x^{x''} \alpha(t)f'(t)dt \geq \alpha(x') \int_{x'}^{x''} f'(t)dt,$$

where the second inequality follows from Lemma 1. So

$$g(x'') - g(x') = \alpha(x')(f(x'') - f(x')) \tag{4}$$

and $g(x'') \geq (>)g(x')$ if $f(x'') \geq (>)f(x')$. QED

Another familiar concept in the theory of monotone comparative statics, and one which is stronger than SCP, is the concept of *increasing differences* (see Milgrom and Shannon (1994) and Topkis (1998)). The function g dominates f by *increasing differences* (see Milgrom and Shannon (1994) and Topkis (1998)) if for any x'' and x' in X , with $x'' > x'$, we have

$$g(x'') - g(x') \geq f(x'') - f(x'). \tag{5}$$

We say that a function g dominates f by *conditional increasing differences* (and denote it by $g \succeq_{IN} f$) if (5) holds for all pairs x' and x'' in X such that $f(x'') \geq f(x)$ for x in $[x', x'']$. Clearly, if g dominates f by conditional increasing differences then

g I-dominates f . The next result gives sufficient conditions under which g and f can be related in this manner.

PROPOSITION 2: *Suppose X is an interval of R and the functions $f, g : X \rightarrow R$ are absolutely continuous on compact intervals in X (and thus f and g are differentiable a.e.). If there is an increasing function $\alpha : X \rightarrow R$ with $\alpha(x) \geq 1$ for all x in X such that $g'(x) \geq \alpha(x)f'(x)$ a.e., then $g \succeq_{\text{IN}} f$.*

Proof: Retrace the proof of Proposition 1; the result follows from (4). QED

Many problems in economic theory involves the maximization of the difference between benefits and costs arising from some activity. The next result shows the relevance of conditional increasing differences to problems of this sort.

PROPOSITION 3: *Let $X \subset R$ and b, \tilde{b} , and c be real-valued functions defined on X , with c an increasing function. If \tilde{b} dominates b by conditional increasing differences, then the function $\tilde{\Pi}$ dominates the function Π by conditional increasing differences, where $\tilde{\Pi}(x) = \tilde{b}(x) - c(x)$ and $\Pi(x) = b(x) - c(x)$. Consequently, $\text{argmax}_{x \in X} \tilde{\Pi}(x) \geq \text{argmax}_{x \in X} \Pi(x)$.*

Proof: Suppose $\Pi(x'') \geq \Pi(x')$ for x in $[x', x'']$. Since c is increasing, we also have $b(x'') \geq b(x')$ for x in $[x', x'']$. By conditional increasing differences $\tilde{b}(x'') - \tilde{b}(x') \geq b(x'') - b(x')$. Adding $-c(x'') + c(x')$ to both sides of this inequality, we obtain

$$\tilde{\Pi}(x'') - \tilde{\Pi}(x') \geq \Pi(x'') - \Pi(x')$$

as required. The final comparative statics statement follows from Theorem 1. QED

There are several issues relating to Proposition 3 that are worth highlighting. Firstly, the proposition does not assume that the benefit functions b and \tilde{b} are increasing in x . It does, of course, require that they be related by conditional increasing differences; furthermore, this cannot be weakened to to SCP or I-dominance (see Milgrom and Shannon (1994)). The assumption that c is increasing is crucial to its proof. Indeed, without this assumption, the conclusion that $\text{argmax}_{x \in X} \tilde{\Pi}(x) \geq \text{argmax}_{x \in X} \Pi(x)$ is *only* possible by assuming that \tilde{b} dominates b by increasing - rather

than conditional increasing - differences (see Athey et al. (1998)).

In this section, we introduced the interval dominance order and showed how it is useful for comparative statics. The theory of comparative statics based on this property admits further development. In particular, we examine in Section 4 how the interval dominance order arises naturally in comparative statics problems under uncertainty. There is also a natural way of developing this theory in a multi-dimensional context. The interval dominance order can be usefully extended to that setting. Notions like quasisupermodularity (Milgrom and Shannon (1994)) and \mathcal{C} -quasisupermodularity (Quah (2007)), which are important for multi-dimensional comparative statics, are essentially variations on the single crossing property; therefore, like the single crossing property, they can also be generalized. We explore these issues in a companion paper (see Quah and Strulovici (2007)).

The next section is devoted to applying the theoretical results obtained so far; readers who are interested in a quick overview of the theory can skip the next section.

3. APPLICATIONS OF THE IDO PROPERTY

Example 2. A very natural application of Propositions 1 and 2 is to the comparative statics of optimal stopping time problems. We consider a simple deterministic problem here; in Quah and Strulovici (2007) we show that the results in the next proposition extend naturally to a stochastic optimal stopping time problem.

Suppose we are interested in maximizing $V_\delta(x) = \int_0^x e^{-\delta t} u(t) dt$ for $x \geq 0$, where $\delta > 0$ and the function $u : R_+ \rightarrow R$ is bounded on compact intervals and measurable. So x may be interpreted as the stopping time, δ is the discount rate, $u(t)$ the cash flow or utility of cash flow at time t (which may be positive or negative), and $V_\delta(x)$ is the discounted sum of the cash flow (or its utility) when x is the stopping time.

We are interested in how the optimal stopping time changes with the discount rate. It seems natural that the optimal stopping time will rise as the discount rate δ falls. This intuition is correct but it cannot be proved by the methods of concave

optimization since V_δ need not be a quasiconcave function. Indeed, it will have a turning point every time u changes sign and its local maxima occur when u changes sign from positive to negative. Changing the discount rate does not change the times at which local maxima are achieved, but it potentially changes the time at which the *global* maximum is achieved, i.e., it changes the optimal stopping time. The next result gives the solution to this problem.

PROPOSITION 4: *Suppose that $\delta > \bar{\delta} > 0$. Then the following holds:*

- (i) $V_{\bar{\delta}} \succeq_{IN} V_\delta$;
- (ii) $\operatorname{argmax}_{x \geq 0} V_{\bar{\delta}}(x) \geq \operatorname{argmax}_{x \geq 0} V_\delta(x)$; and
- (iii) $\max_{x \geq 0} V_{\bar{\delta}}(x) \geq \max_{x \geq 0} V_\delta(x)$.

Proof: The functions $V_{\bar{\delta}}$ and V_δ are absolutely continuous and thus differentiable a.e.; moreover,

$$V_{\bar{\delta}}'(x) = e^{-\bar{\delta}x} u(x) = e^{(\delta - \bar{\delta})x} V_\delta'(x).$$

Note that the function $\alpha(x) = e^{(\delta - \bar{\delta})x}$ is increasing and greater than 1. So part (i) follows from Proposition 2 and part (ii) from Theorem 1. For (iii), let us suppose that $V_\delta(x)$ is maximized at $x = x^*$. Then for all x in $[0, x^*]$, $V_\delta(x) \leq V_\delta(x^*)$. Since $V_{\bar{\delta}}(0) = V_\delta(0) = 0$, the fact that $V_{\bar{\delta}} \succeq_{IN} V_\delta$ now guarantees that $V_{\bar{\delta}}(x^*) \geq V_\delta(x^*)$. Finally, note that $\max_{x \geq 0} V_{\bar{\delta}}(x) \geq V_{\bar{\delta}}(x^*)$. QED

Arrow and Levhari (1969) has a version of Proposition 4(iii) (but not (i) and (ii)). They require π to be a continuous function; with this assumption, they show that the value function \bar{V} , defined by $\bar{V}(\delta) = \max_{x \geq 0} V_\delta(x)$ is right differentiable and has a negative derivative. This result is the crucial step (in their proof) guaranteeing the existence of a *unique* internal rate of return for an investment project, i.e., a unique δ^* such that $\bar{V}(\delta^*) = 0$. It is possible for us to extend and apply Proposition 4 to prove something along these lines, but we shall not do so in this paper.⁷

Example 3. Consider a firm that chooses output x to maximize profit, given by $\Pi(x) = xP(x) - C(x)$, where P is the inverse demand function and C is the cost

⁷We would like to thank H. Polemarchakis for pointing out Arrow and Levhari's result.

function. Imagine that there is a change in market conditions, so that the both P and C are changed to \tilde{P} and \tilde{C} respectively. When can we say that $\operatorname{argmax}_{x \geq 0} \tilde{\Pi}(x) \geq \operatorname{argmax}_{x \geq 0} \Pi(x)$? By Theorem 1, this holds if $\tilde{\Pi}$ I-dominates Π . Intuitively, we will expect this to hold if the increase in the inverse demand is greater than any increase in costs. This idea can be formalized in the following manner.

Assume that all the functions are differentiable, that \tilde{P} and P take strictly positive values, and that the cost functions are strictly increasing. Define $a(x) = \tilde{\Pi}(x)/\Pi(x)$. Then

$$\begin{aligned} \tilde{\Pi}'(x) &= a'(x)xP(x) + a(x)(xP(x))' - \tilde{C}'(x) \\ &\geq a(x)(xP(x))' - \left[\frac{\tilde{C}'(x)}{C'(x)} \right] C'(x), \end{aligned}$$

where the inequality follows since $xP(x) > 0$. Now suppose we make we assume that

$$a(x) = \frac{\tilde{P}(x)}{P(x)} \geq \frac{\tilde{C}'(x)}{C'(x)}, \quad (6)$$

then we obtain $\tilde{\Pi}'(x) \geq a(x)\Pi'(x)$. By Proposition 1, $\tilde{\Pi}$ I-dominates Π if a is increasing and (6) holds; in other words, the ratio of the inverse demand functions is increasing in x and greater than the ratio of the marginal costs.⁸

Example 4. We wish to show, in the context of a standard optimal growth model, that lowering the discount rate of the representative agent leads to capital deepening: specifically, a higher capital stock at all times. Formally, the agent solves

$$\max U(c, k) = \int_0^\infty e^{-\delta s} u(c(s), k(s), s) ds \quad \text{subject to}$$

(a) $\dot{k}(t) = H(c(t), k(t), t)$; (b) $k(t) \geq 0$ and $0 \leq c(t) \leq Q(k(t), t)$; and (c) $k(0) = k_0$.

The scalars $c(t)$ and $k(t)$ are the consumption and capital stock at time t respectively. Q is the production function, u the felicity function, and δ the discount rate. It is standard to have $H(c(t), k(t), t) = Q(k(t), t) - c(t) - \eta k(t)$ where $\eta \in (0, 1)$ is the

⁸Note that our argument does not require that P be decreasing in x .

(constant) rate of depreciation, but our result does not rely on this functional form. We also allow felicity to depend on consumption and capital (rather than just the former) and both the felicity and production functions may vary with t directly. In these respects, our model specification is more general than what is often assumed.

Capital is said to be *beneficial* if the optimal value of $\int_{t_1}^{t_2} e^{-\delta s} u(c(s), k(s), s) ds$ subject to (a), (b), and the boundary conditions $k(t_1) = k_1$ and $k(t_2) = k_2$ is strictly increasing in k_1 . In other words, raising the capital stock at t_1 strictly increases the utility achieved in the period $[t_1, t_2]$. Clearly, this is a mild condition which, in essence, is guaranteed if felicity strictly increases with consumption and production strictly increases with capital. This condition is all that is needed to guarantee that a lower discount rate leads to capital deepening.⁹

PROPOSITION 5: *Suppose that capital is beneficial and (\bar{c}, \bar{k}) and (\hat{c}, \hat{k}) are solutions to the optimal growth problem at discount rates $\bar{\delta}$ and $\hat{\delta}$ respectively. If $\hat{\delta} < \bar{\delta}$, then $\hat{k}(t) \geq \bar{k}(t)$ for all $t \geq 0$.*

Proof: First, observe that the function $F(\cdot, T, \hat{\delta})$ I-dominates $F(\cdot, T, \bar{\delta})$, where

$$F(t, T, \delta) \equiv \int_t^T e^{-\delta s} \left[u(\hat{c}(s), \hat{k}(s), s) - u(\bar{c}(s), \bar{k}(s), s) \right] ds \text{ with } \delta = \bar{\delta}, \hat{\delta}.$$

This follows from Proposition 2 since $F_t(t, T, \hat{\delta}) = e^{(\bar{\delta} - \hat{\delta})t} F_t(t, T, \bar{\delta})$. In particular, suppose that for all t in some interval $[\underline{t}, T]$ we have $F(t, T, \bar{\delta}) \leq 0$. Since $F(T, T, \bar{\delta}) = 0$, we obtain by the I-dominance property that $F(\underline{t}, T, \hat{\delta}) \leq F(T, T, \hat{\delta}) = 0$. Indeed, we can say more. If $F(t, T, \bar{\delta}) < 0$ for t in a set of non-zero measure in $[\underline{t}, T]$, then $F(\underline{t}, T, \hat{\delta}) < 0$.¹⁰

⁹The reader may consult Boyd and Becker (1997) for a discussion of other results (in particular, Amir (1996)) on the response of capital to the discount rate. Typically, these results are valid for multiple capital goods and require, amongst other things, the supermodularity of the value function. When specialized to the case of one capital good, those assumptions are still stronger than the ones made in Proposition 5.

¹⁰We can see this by re-tracing the proof of Lemma 1. In Lemma 1, the inequality (3) is strict if (a) $\int_x^{x''} h(t) dt > 0$ for x in a set with positive measure in (x', x'') and (b) α is a strictly increasing function. Note that in this application, $\alpha(t) = e^{(\bar{\delta} - \hat{\delta})t}$, which is strictly increasing in t .

Suppose that, contrary to our claim, there is some time t' at which $\bar{k}(t') > \hat{k}(t')$. Let \underline{t} be the largest t below \underline{t} such that $\bar{k}(\underline{t}) = \hat{k}(\underline{t})$; the existence of \underline{t} is guaranteed by the continuity of \bar{k} and \hat{k} and the fact that $\bar{k}(0) = \hat{k}(0)$. Let T be the earliest time after t' at which $\bar{k}(T) = \hat{k}(T)$. Set $T = \infty$ if no such time exists. So for t in the interval $[\underline{t}, T]$, we have $\bar{k}(t) \geq \hat{k}(t)$, with a strict inequality for t in (\underline{t}, T) . For such a t , denote by (\tilde{c}, \tilde{k}) the path that maximizes $\int_t^T e^{-\bar{\delta}s} u(c(s), k(s), s) ds$ subject to (a), (b), and the boundary conditions $k(t) = \hat{k}(t)$ and $k(T) = \bar{k}(T)$. We have

$$\int_t^T e^{-\bar{\delta}s} u(\bar{c}(s), \bar{k}(s), s) ds \geq \int_t^T e^{-\bar{\delta}s} u(\tilde{c}(s), \tilde{k}(s), s) ds \geq \int_t^T e^{-\hat{\delta}s} u(\hat{c}(s), \hat{k}(s), s) ds.$$

The second inequality follows from the optimality of (\tilde{c}, \tilde{k}) . The first inequality is an equality if $t = \underline{t}$ and (by the fact that capital is beneficial) a strict inequality if t is in (\underline{t}, T) . So $F(t, T, \bar{\delta}) \leq 0$ for t in $[\underline{t}, T]$, with a strict inequality for t in (\underline{t}, T) . We have shown in the preceding paragraph that this implies that $F(\underline{t}, T, \hat{\delta}) < 0$. This is a contradiction because the optimality of (\hat{c}, \hat{k}) at the discount rate $\hat{\delta}$ means that it cannot accumulate strictly less utility over the interval $[\underline{t}, T]$ than (\bar{c}, \bar{k}) . QED

4. THE INTERVAL DOMINANCE ORDER WHEN THE STATE IS UNCERTAIN

Consider the following problem. Let $\{f(\cdot, s)\}_{s \in S}$ be a family of functions parameterized by s in S , an interval of R , with each function $f(\cdot, s)$ mapping Y , an interval of R , to R . Assume that all the functions are quasiconcave, with their peaks increasing in s ; by this we mean that $\operatorname{argmax}_{x \in Y} f(x, s'') \geq \operatorname{argmax}_{x \in Y} f(x, s')$ whenever $s'' > s'$. (Note that since each function $f(\cdot, s)$ is quasiconcave, it either has a unique maximizer or they must form an interval.) We shall refer to such a family of functions as a QCIP family, where QCIP stands for *quasiconcave with increasing peaks*. Interpreting s to be the state of the world, an agent has to choose x under uncertainty, i.e., before s is realized. We assume the agent maximizes the expected value of his objective; formally, he maximizes

$$F(x, \lambda) = \int_{s \in S} f(x, s) \lambda(s) ds,$$

where $\lambda : S \rightarrow R$ is the density function defined over the states of the world. It is natural to think that if the agent considers the higher states to be more likely, then his optimal value of x will increase. Is this true? More generally, we can ask the same question if the functions $\{f(\cdot, s)\}_{s \in S}$ form an IDO family, i.e., a family of regular functions $f(\cdot, s) : X \rightarrow R$, with $X \subseteq R$, such that $f(\cdot, s'')$ I-dominates $f(\cdot, s')$ whenever $s'' > s'$.

One way of formalizing the notion that higher states are more likely is via the monotone likelihood ratio (MLR) property. Let λ and γ be two density functions defined on the interval S of R and assume that $\lambda(s) > 0$ for s in S . We call γ an *MLR shift* of λ if $\gamma(s)/\lambda(s)$ is increasing in s . For density changes of this kind, there are two results that come close, though not quite, to addressing the problem we posed.

Ormiston and Schlee (1993) identify some conditions under which an MLR shift in the density function will raise the agent's optimal choice. Amongst other conditions, they assume that $F(\cdot; \lambda)$ is quasiconcave. This will hold if all the functions in the family $\{f(\cdot, s)\}_{s \in S}$ are concave but will not generally hold if the functions are just quasiconcave. Athey (2002) has a related result which says that an MLR shift will lead to a higher optimal choice of x provided $\{f(\cdot, s)\}_{s \in S}$ is an SCP family. As we had already pointed out in Example 1, a QCIP family need not be an SCP family.

The next result gives the solution to the problem we posed.

THEOREM 2: *Let S be an interval of R and $\{f(\cdot, s)\}_{s \in S}$ be an IDO family. Then $F(\cdot, \gamma) \succeq_I F(\cdot, \lambda)$ if γ is an MLR shift of λ . Consequently, $\operatorname{argmax}_{x \in X} F(x, \gamma) \geq \operatorname{argmax}_{x \in X} F(x, \lambda)$.*

Notice that since $\{f(\cdot, s)\}_{s \in S}$ in Theorem 2 is assumed to be an IDO family, we know (from Theorem 1) that $\operatorname{argmax}_{x \in X} f(x, s'') \geq \operatorname{argmax}_{x \in X} f(x, s')$. Thus Theorem 2 guarantees that the comparative statics which holds when s is known also holds when s is unknown but experiences an MLR shift.¹¹

The proof of Theorem 2 requires a lemma (stated below). Its motivation arises

¹¹We are echoing an observation that was also made by Athey (2002) in a similar context.

from the observation that if $g \succeq_{sc} f$, then for any $x'' > x'$ such that $g(x') - g(x'') \geq (>) 0$, we must also have $f(x') - f(x'') \geq (>) 0$. Lemma 2 is the (less trivial) analog of this observation in the case when $g \succeq_I f$.

LEMMA 2: *Let X be a subset of R and f and g two regular functions defined on X . Then $g \succeq_I f$ if and only if the following property holds:*

(M) if $g(x') \geq g(x)$ for x in $[x', x'']$ then

$$g(x') - g(x'') \geq (>) 0 \implies f(x') - f(x'') \geq (>) 0.$$

Proof: Suppose $x' < x''$ and $g(x') \geq g(x)$ for x in $[x', x'']$. There are two possible ways for property (M) to be violated. One possibility is that $f(x'') > f(x')$. By regularity, we know that $\operatorname{argmax}_{x \in [x', x'']} f(x)$ is nonempty; choosing x^* in this set, we have $f(x^*) \geq f(x)$ for all x in $[x', x^*]$, with $f(x^*) \geq f(x'') > f(x')$. Since $g \succeq_I f$, we must have $g(x^*) > g(x')$, which is a contradiction.

The other possible violation of (M) occurs if $g(x') > g(x'')$ but $f(x') = f(x'')$. By regularity, we know that $\operatorname{argmax}_{x \in [x', x'']} f(x)$ is nonempty, and if f is maximized at x^* with $f(x^*) > f(x')$, then we are back to the case considered above. So assume that x' and x'' are both in $\operatorname{argmax}_{x \in [x', x'']} f(x)$. Since $f \succeq_I g$, we must have $g(x'') \geq g(x')$, contradicting our initial assumption.

So we have shown that (M) holds if $g \succeq_I f$. The proof that (M) implies $g \succeq_I f$ is similar. QED

Proof of Theorem 2: This consists of two parts. Firstly, we prove that if $F(x'', \lambda) \geq F(x, \lambda)$ for all x in $[x', x'']$, then, for any \tilde{s} in S ,

$$\int_{\tilde{s}}^{s^*} (f(x'', s) - f(x', s)) \lambda(s) ds \geq 0 \tag{7}$$

(where s^* denotes the supremum of S). Assume instead that there is \bar{s} such that

$$\int_{\bar{s}}^{s^*} (f(x'', s) - f(x', s)) \lambda(s) ds < 0. \tag{8}$$

By the regularity of $f(\cdot, \bar{s})$, there is \bar{x} that maximizes f in $[x', x'']$. In particular, $f(\bar{x}, \bar{s}) \geq f(x, \bar{s})$ for all x in $[\bar{x}, x'']$. Since $\{f(\cdot, s)\}_{s \in S}$ is an IDO family of regular functions, we also have $f(\bar{x}, s) \geq f(x'', s)$ for all $s \leq \bar{s}$ (using Lemma 2). Thus

$$\int_{s_*}^{\bar{s}} (f(\bar{x}, s) - f(x'', s)) \lambda(s) ds \geq 0, \quad (9)$$

where s_* is the infimum of S . Notice also that $f(\bar{x}, \bar{s}) \geq f(x, \bar{s})$ for all x in $[x', \bar{x}]$, which implies that $f(\bar{x}, s) \geq f(x', s)$ for all $s \geq \bar{s}$. Aggregating across s we obtain

$$\int_{\bar{s}}^{s^*} (f(\bar{x}, s) - f(x', s)) \lambda(s) ds \geq 0. \quad (10)$$

It follows from (8) and (10) that

$$\begin{aligned} \int_{\bar{s}}^{s^*} (f(\bar{x}, s) - f(x'', s)) \lambda(s) ds &= \int_{\bar{s}}^{s^*} (f(\bar{x}, s) - f(x', s)) \lambda(s) ds \\ &\quad + \int_{\bar{s}}^{s^*} (f(x', s) - f(x'', s)) \lambda(s) ds \\ &> 0. \end{aligned}$$

Combining this with (9), we obtain

$$\int_{s_*}^{s^*} (f(\bar{x}, s) - f(x'', s)) \lambda(s) ds > 0;$$

in other words, $F(\bar{x}, \lambda) > F(x'', \lambda)$ which is a contradiction.

Given (7), the function $H(\cdot, \lambda) : [s_*, s^*] \rightarrow R$ defined by

$$H(\tilde{s}, \lambda) = \int_{s_*}^{\tilde{s}} (f(x'', s) - f(x', s)) \gamma(s) ds$$

satisfies $H(s^*, \lambda) \geq H(\tilde{s}, \lambda)$ for all \tilde{s} in $[s_*, s^*]$. Defining $H(\cdot, \gamma)$ in an analogous fashion, we also have $H'(\tilde{s}, \gamma) = [\gamma(s)/\lambda(s)]H'(\tilde{s}, \lambda)$ for \tilde{s} in S . Since γ is an upward MLR shift of λ , the ratio $\gamma(s)/\lambda(s)$ is increasing in s . By Proposition 1, $H(\cdot, \gamma) \succeq_I H(\cdot, \lambda)$. In particular, we have $H(s^*, \gamma) \geq (>)H(s_*, \gamma) = 0$ if $H(s^*, \lambda) \geq (>)H(s_*, \lambda) = 0$. Re-writing this, we have $F(x'', \gamma) \geq (>)F(x', \gamma)$ if $F(x'', \lambda) \geq (>)F(x', \lambda)$. QED

Note that Theorem 2 remains true if S is not an interval; in Appendix A, we prove Theorem 2 in the case where S is a finite set of states.

We turn now to two applications of Theorem 2.

Example 1 continued. Recall that in state s , the firm's profit is $\Pi(x, s)$. It achieves its maximum at $x^*(s) = s$, with $\Pi(s, s) = (1 - c(s) - D)s$, which is strictly positive by assumption. The firm has to choose its capacity before the state of the world is realized; we assume that s is drawn from S , an interval in R , and has a distribution given by the density function $\lambda : S \rightarrow R$. We can think of the firm as maximizing its expected profit, which is $\int_S \Pi(x, s)\lambda(s)ds$, or more generally, let us assume that it maximizes the expected utility from profit, i.e., it maximizes $U(x, \lambda) = \int_S u(\Pi(x, s), s)\lambda(s)ds$, where, for each s , the function $u(\cdot, s) : R \rightarrow R$ is strictly increasing. The family $\{u(\Pi(\cdot, s), s)\}_{s \in S}$ consists of quasiconcave functions, the peaks of which are increasing in s . This is an IDO family. By Theorem 2, we know that an upward MLR shift of the density function will lead the firm to choose a greater capacity.

Example 5. Consider a firm that has to decide on when to launch a new product. The more time the firm gives itself, the more it can improve the quality of the product and its manufacturing process, but it also knows that there is a rival about to launch a similar product. In formal terms, we assume that the firm's profit (if it is not anticipated by its rival) is an increasing function of time $\bar{\pi} : R_+ \rightarrow R_+$. If the rival launches its product at time s , then the firm's profit falls to $w(s)$ (in R). In other words, the firm's profit in state s is $\pi(t, s) = \bar{\pi}(t)$ for $t \leq s$ and $w(s)$ for $t > s$, where $w(s) < \bar{\pi}(s)$. Clearly, each $\pi(\cdot, s)$ is a quasiconcave function and $\{\pi(\cdot, s)\}_{s \in S}$ is an IDO family. The firm decides on the launch date t by maximizing $F(t, \lambda) = \int_{s \in S} \pi(t, s)\lambda(s)ds$, where $\lambda : R_+ \rightarrow R$ is the density function over s . By Theorem 2, if the firm thinks that it is less likely that the rival will launch early, in the sense that there is an MLR shift in the density function, then it will decide on a later launch date.

Note that we impose no restrictions on the function $w : R_+ \rightarrow R$, which gives the firm's profit should it be anticipated by its rival at time s . If w is an increasing

function of s , then one can check that $\{\pi(\cdot, s)\}_{s \in S}$ is an SCP family, but Theorem 2 gives us the desired conclusion without making this stronger assumption.

5. COMPARING INFORMATION STRUCTURES¹²

Consider an agent who, as in the previous section, has to make a decision before the state of the world (s) is realized, where the set of possible states S is a subset of R . Suppose that, before he makes his decision, the agent observes a signal z . This signal is potentially informative of the true state of the world; we refer to the collection $\{H(\cdot|s)\}_{s \in S}$, where $H(\cdot|s)$ is the distribution of the signal z conditional on s , as the *information structure* of the decision maker's problem. (Whenever convenient, we shall simply call this information structure H .) We assume that, for every s , $H(\cdot|s)$ admits a density function and has the compact interval Z as its support. We say that H is *MLR-ordered* if $H(\cdot|s'')$ is an MLR shift of $H(\cdot|s')$ whenever $s'' > s'$.

We assume that the agent has a prior distribution λ on S . We allow either of the following: (i) S is a compact interval and λ admits a density function with S as its support or (ii) S is finite and λ has S as its support.

The agent's *decision rule* (under H) is a map from Z to set of actions X (contained in R). We denote his posterior distribution (on S) upon observing z by λ_H^z ; so the agent with a decision rule $\phi : Z \rightarrow X$ will have an ex ante utility given by

$$\mathcal{U}(\phi, H, \lambda) = \int_{z \in Z} \left[\int_{s \in S} u(\phi(z), s) d\lambda_H^z \right] dM_{H, \lambda} = \int_{Z \times S} u(\phi(z), s) dJ_{H, \lambda}$$

where $M_{H, \lambda}$ is the marginal distribution of z and $J_{H, \lambda}$ the joint distribution of (z, s) given H and λ . A decision rule $\hat{\phi} : Z \rightarrow X$ that maximizes the agent's (posterior) expected utility at each realized signal is called an *H-optimal* decision rule. We assume that X is compact and that, for all s , the function $u(\cdot, s)$ is continuous. This guarantees that $\hat{\phi}$ exists. The agent's ex ante utility using such a rule is denoted by $\mathcal{V}(H, \lambda, u)$.

¹²We are very grateful to Ian Jewitt for introducing us to the literature in this section and the next and for extensive discussions.

Consider now an alternative information structure given by the collection $\{G(\cdot|s)\}_{s \in S}$; we assume that $G(\cdot|s)$ admits a density function and has the compact interval Z as its support. What conditions will guarantee that the information structure H is more favorable than G in the sense of offering the agent a higher ex ante utility; in other words, how can we guarantee that $\mathcal{V}(H, \lambda, u) \geq \mathcal{V}(G, \lambda, u)$?

It is well known that this holds if H is *more informative* than G according to the criterion developed by Blackwell (1953); furthermore, this criterion is also necessary if one does not impose significant restrictions on u (see Blackwell (1953) or, for a recent textbook treatment, Gollier (2001)). We wish instead to consider the case where a significant restriction *is* imposed on u ; specifically, we assume that $\{u(\cdot, s)\}_{s \in S}$ is an IDO family. We show that, in this context, a different notion of informativeness due to Lehmann (1988) is the appropriate concept.¹³

Our assumptions on H guarantee that, for any s , $H(\cdot|s)$ admits a density function with support Z ; therefore, for any (z, s) in $Z \times S$, there exists a unique element in Z , which we denote by $T(z, s)$, such that $H(T(z, s)|s) = G(z|s)$. We say that H is *more accurate* than G if T is an increasing function of s .¹⁴ Our goal in this section is to prove the following result.

THEOREM 3: *Suppose $\{u(\cdot, s)\}_{s \in S}$ is an IDO family, G is MLR-ordered, and λ is the agent's prior distribution on S . If H is more accurate than G , we obtain*

$$\mathcal{V}(H, \lambda, u) \geq \mathcal{V}(G, \lambda, u). \quad (11)$$

¹³Jewitt (2006) gives the precise sense in which Lehmann's concept is weaker than Blackwell's (see also Lehmann (1988) and Persico (2000)) and also discusses its relationship with the concept of *concordance*. Some papers with economic applications of Lehmann's concept of informativeness are Persico (2000), Athey and Levin (2001), Levin (2001), Bergemann and Valimaki (2002), and Jewitt (2006). Athey and Levin (2001) explore other related concepts of informativeness and their relationship with the payoff functions. The manner in which these papers are related to Lehmann's (1988) result and to each other is not straightforward; Jewitt (2006) provides an overview.

¹⁴The concept is Lehmann's; the term *accuracy* follows Persico (2000).

This theorem generalizes a number of earlier results. Lehmann (1988) establishes a special case of Theorem 3 in which $\{u(\cdot, s)\}_{s \in S}$ is a QCIP family. Persico (1996) has a version of Theorem 3 in which $\{u(\cdot, s)\}_{s \in S}$ is an SCP family, but he requires the optimal decision rule to vary smoothly with the signal, a property that is not generally true without the *sufficiency* of the first order conditions for optimality. Jewitt (2006) proves Theorem 3 for the general SCP case.¹⁵

To prove Theorem 3, we first note that if G is MLR-ordered, then the family of posterior distributions $\{\lambda_H^z\}_{z \in Z}$ is also MLR-ordered, i.e., if $z'' > z'$ then $\lambda_H^{z''}$ is an MLR shift of $\lambda_H^{z'}$.¹⁶ Since $\{u(\cdot, s)\}_{s \in S}$ is an IDO family, Theorem 2 guarantees that the G -optimal decision rule can be chosen to be increasing with z . Therefore, Theorem 3 is valid if we can show that for *any* increasing decision rule $\psi : Z \rightarrow X$ under G there is a rule $\phi : Z \rightarrow X$ under H that gives a higher ex ante utility, i.e.,

$$\int_{Z \times S} u(\phi(z), s) dJ_{H, \lambda} \geq \int_{Z \times S} u(\psi(z), s) dJ_{G, \lambda}.$$

This inequality in turn follows from aggregating (across s) the inequality (12) below.

PROPOSITION 6: *Suppose $\{u(\cdot, s)\}_{s \in S}$ is an IDO family and H is more accurate than G . Then for any increasing decision rule $\psi : Z \rightarrow X$ under G , there is an increasing decision rule $\phi : Z \rightarrow X$ under H such that, at each state s , the distribution of utility induced by ϕ and $H(\cdot|s)$ first order stochastically dominates the distribution of utility induced by ψ and $G(\cdot|s)$. Consequently, at each state s ,*

$$\int_{z \in Z} u(\phi(z), s) dH(z|s) \geq \int_{z \in Z} u(\psi(z), s) dG(z|s). \quad (12)$$

(At a given state s , a decision rule ρ and a distribution on z induces a distribution of utility in the following sense: for any measurable set U of R , the probability of

¹⁵However, there is a sense in which it is incorrect to say that Theorem 3 generalizes Lehmann's result. The criterion employed by us here (and indeed by Persico (1996) and Jewitt (2006) as well) - comparing information structures with the ex ante utility - is less stringent than the criterion Lehmann used. In the next section we shall compare information structures using precisely the same criterion as Lehmann and prove a result (Corollary 1) that is stronger than Theorem 3.

¹⁶This is not hard to prove; indeed, the two properties are equivalent.

$\{u \in U\}$ equals the probability of $\{z \in Z : u(\rho(z), s) \in U\}$. So it is meaningful to refer, as this proposition does, to the distribution of utility at each s .)

Our proof of Proposition 6 requires the following lemma.

LEMMA 3: *Suppose $\{u(\cdot, s)\}_{s \in S}$ is an IDO family and H is more accurate than G . Then for any increasing decision rule $\psi : Z \rightarrow X$ under G , there is an increasing decision rule $\phi : Z \rightarrow X$ under H such that, for all (z, s) ,*

$$u(\phi(T(z, s)), s) \geq u(\psi(z), s). \quad (13)$$

Proof: We shall only demonstrate here how we construct ϕ from ψ in the case where ψ takes only finitely many values. This is true, in particular, when the set of actions X is finite. The extension to the case where the range of ψ is infinite is shown in Appendix B; the proof that ϕ is increasing is also postponed to that appendix. The proof below assumes that S is a compact interval, but it can be modified in an obvious way to deal with the case where S is finite.

For every \bar{t} in Z and s in S , there is a unique \bar{z} in Z such that $\bar{t} = T(\bar{z}, s)$. This follows from the fact that $G(\cdot|s)$ is a strictly increasing continuous function (since it admits a density function with support Z). We write $\bar{z} = \tau(\bar{t}, s)$; note that because T is increasing in both its arguments, the function $\tau : Z \times S \rightarrow Z$ is decreasing in s . Note also that (13) is equivalent to

$$u(\phi(t), s) \geq u(\psi(\tau(t, s)), s). \quad (14)$$

We will now show how $\phi(t)$ may be chosen to satisfy (14).

Note that because τ is decreasing and ψ is increasing the function $\psi(\tau(t, \cdot))$ is decreasing (in s). This fact, together with our assumption that ψ takes finitely many values, allow us to partition $S = [s^*, s^{**}]$ into the sets S_1, S_2, \dots, S_M , where M is odd, with the following properties: (i) if $m > n$, then any element in S_m is greater than any element in S_n ; (ii) whenever m is odd, S_m is a singleton, with $S_1 = \{s^*\}$ and

$S_M = \{s^{**}\}$; (iii) when m is even, S_m is an open interval; (iv) for any s' and s'' in S_m , we have $\psi(\tau(t, s')) = \psi(\tau(t, s''))$; and (v) for s'' in S_m and s' in S_n such that $m > n$, $\psi(\tau(t, s')) \geq \psi(\tau(t, s''))$.

In other words, we have partitioned S into finitely many sets, so that within each set, $\psi(\tau(t, \cdot))$ takes the same value. Denoting $\psi(\tau(t, s))$ for s in S_m by ψ_m , (v) says that $\psi_1 \geq \psi_2 \geq \psi_3 \geq \dots \geq \psi_M$. Establishing (14) involves finding $\phi(t)$ such that

$$u(\phi(t), s_m) \geq u(\psi_m, s_m) \text{ for any } s_m \in S_m; m = 1, 2, \dots, M. \quad (15)$$

In the interval $[\psi_2, \psi_1]$, we pick the largest action $\hat{\phi}_2$ that maximizes $u(\cdot, s^*)$ in that interval. This exists because $u(\cdot, s^*)$ is continuous and $X \cap [\psi_2, \psi_1]$ is compact. By the IDO property,

$$u(\hat{\phi}_2, s_m) \geq u(\psi_m, s_m) \text{ for any } s_m \in S_m; m = 1, 2. \quad (16)$$

Recall that S_3 is a singleton; we call that element s_3 . The action $\hat{\phi}_4$ is chosen to be the largest action in the interval $[\psi_4, \hat{\phi}_2]$ that maximizes $u(\cdot, s_3)$. Since ψ_3 is in that interval, we have $u(\hat{\phi}_4, s_3) \geq u(\psi_3, s_3)$. Since $u(\hat{\phi}_4, s_3) \geq u(\psi_4, s_3)$, the IDO property guarantees that $u(\hat{\phi}_4, s_4) \geq u(\psi_4, s_4)$ for any s_4 in S_4 . Using the IDO property again (specifically, Lemma 2), we have $u(\hat{\phi}_4, s_m) \geq u(\hat{\phi}_2, s_m)$ for s_m in S_m ($m = 1, 2$) since $u(\hat{\phi}_4, s_3) \geq u(\hat{\phi}_2, s_3)$. Combining this with (16), we have found $\hat{\phi}_4$ in $[\psi_4, \hat{\phi}_2]$ such that

$$u(\hat{\phi}_4, s_m) \geq u(\psi_m, s_m) \text{ for any } s_m \in S_m; m = 1, 2, 3, 4. \quad (17)$$

We can repeat the procedure finitely many times, at each stage choosing $\hat{\phi}_{m+1}$ (for m odd) as the largest element maximizing $u(\cdot, s_m)$ in the interval $[\psi_{m+1}, \hat{\phi}_{m-1}]$, and finally, choosing $\hat{\phi}_{M+1}$ as the largest element maximizing $u(\cdot, s^{**})$ in the interval $[\psi_M, \hat{\phi}_{M-1}]$. It is clear that $\phi(t) = \hat{\phi}_M$ will satisfy (15). QED

Proof of Proposition 6: Let \tilde{z} denote the random signal received under information structure G and let \tilde{u}^G denote the (random) utility achieved when the decision rule ψ is used. Correspondingly, we denote the random signal received under H by \tilde{t} , with

\tilde{u}^H denoting the utility achieved by the rule ϕ , as constructed in Lemma 3. Observe that for any fixed utility level u' and at a given state s' ,

$$\begin{aligned}
\Pr[\tilde{u}^H \leq u' | s = s'] &= \Pr[u(\phi(\tilde{t}), s') \leq u' | s = s'] \\
&= \Pr[u(\phi(T(\tilde{z}, s')), s') \leq u' | s = s'] \\
&\leq \Pr[u(\psi(\tilde{z}), s') \leq u' | s = s'] \\
&= \Pr[\tilde{u}^G \leq u' | s = s']
\end{aligned}$$

where the second equality comes from the fact that, conditional on $s = s'$, the distribution of \tilde{t} coincides with that of $T(\tilde{z}, s')$, and the inequality comes from the fact that $u(\phi(T(z, s')), s') \geq u(\psi(z), s')$ for all z (by Lemma 3).

Finally, the fact that, given the state, the conditional distribution of \tilde{u}^H first order stochastically dominates \tilde{u}^G means that the conditional mean of \tilde{u}^H must also be higher than that of \tilde{u}^G . QED

Example 1 continued. As a simple application of Theorem 3, we return again to this example (previously discussed in Sections 2 and 4), where a firm has to decide on its production capacity before the state of the world is realized. Recall that the profit functions $\{\Pi(\cdot, s)\}_{s \in S}$ form an IDO (though not necessarily SCP) family. Suppose that before it makes its decision, the firm receives a signal z from the information structure G . Provided G is MLR-ordered, we know that the posterior distributions (on S) will also be MLR-ordered (in z). It follows from Theorem 2 that a higher signal will cause the firm to decide on a higher capacity. Assuming the firm is risk neutral, its ex ante expected profit is $\mathcal{V}(\lambda, G, \Pi)$, where λ is the firm's prior on S . Theorem 3 tells us that a more accurate information structure H will lead to a higher ex ante expected profit; the difference $\mathcal{V}(\lambda, H, \Pi) - \mathcal{V}(\lambda, G, \Pi)$ represents what the firm is willing to spend for the more accurate information structure.

It is clear that we can extend our analysis of Example 5 (in Section 4) in a similar way, i.e., we can introduce and compare information structures available to the firm for deciding the date of its product launch.

It is worth pointing out that our use of Proposition 6 to prove Theorem 3 (via (12)) does not fully exploit the property of first order stochastic dominance that Proposition 5 obtains. Our next application is one where this stronger conclusion is crucial.

Example 6. There are N investors, with investor i having wealth $w_i > 0$ and the strictly increasing Bernoulli utility function v_i . These investors place their wealth with a manager who has to decide on an investment policy; specifically, the manager must allocate the total pool of funds $W = \sum_{i=1}^N w_i$ between a risky asset, with return s in state s , and a safe asset with return $r > 0$. Denoting the fraction invested in the risky asset by x , investor i 's utility (as a function of x and s) is given by $u_i(x, s) = v_i((xs + (1-x)r)w_i)$. It is easy to see that $\{u_i(\cdot, s)\}_{s \in S}$ is an IDO family. Indeed, it is also an SCP and a QCIP family: for $s > r$, $u_i(\cdot, s)$ is strictly increasing in x ; for $s = r$, $u_i(\cdot, r)$ is the constant $v_i(rw_i)$; and for $s < r$, $u_i(\cdot, s)$ is strictly decreasing in x .

Before she makes her portfolio decision, the manager receives a signal z from some information structure G . She employs the decision rule ψ , where $\psi(z)$ (in $[0, 1]$) is the fraction of W invested in the risky asset. We assume that ψ is increasing in the signal. (We shall justify this assumption in the next section.) Suppose that the manager now has access to a superior information structure H . By Proposition 6, there is an increasing decision rule ϕ under H such that, at any state s , the distribution of investor k 's utility under H and ϕ first order stochastically dominates the distribution of k 's utility under G and ψ . In particular, (12) holds for $u = u_k$. Aggregating across states we obtain $\mathcal{U}_k(\phi, H, \lambda_k) \geq \mathcal{U}_k(\psi, G, \lambda_k)$, where λ_k is investor k 's (subjective) prior; in other words, k 's ex ante utility is higher with the new information structure and the new decision rule.

But even more can be said because, for any other investor i , $u_i(\cdot, s)$ is a strictly increasing transformation of $u_k(\cdot, s)$, i.e., there is a strictly increasing function f such that $u_i = f \circ u_k$. It follows from Proposition 6 that (12) is true, not just for $u = u_k$ but for $u = u_i$. Aggregating across states, we obtain $\mathcal{U}_i(\phi, H, \lambda_i) \geq \mathcal{U}_i(\psi, G, \lambda_i)$, where λ_i

is investor i 's prior.

To summarize, we have shown the following: though different investors may have different attitudes towards risk aversion and different priors, the greater accuracy of H compared to G allows the manager to implement a new decision rule that gives greater ex ante utility to *every* investor.

Finally, we turn to the following question: how important is the accuracy criterion to the results in this section? For example, we may wonder if the conclusion in Theorem 3 is, in a sense, too strong. Theorem 3 tells us that when H is more accurate than G , it gives the agent a higher ex ante utility for any prior that he may have on S . This raises the possibility that the accuracy criterion may be weakened if we only wish H to give a higher ex ante utility than G for a *particular* prior. However, this is not the case, as the next result shows.

PROPOSITION 7: *Let S be finite, and H and G two information structures on S . If (11) holds at a given prior λ^* which has S as its support and for any SCP family $\{u(\cdot, s)\}_{s \in S}$, then (11) holds at any prior λ which has S as its support and for any SCP family.*

Proof: Given a prior λ with S as its support, and given the SCP family $\{u(\cdot, s)\}_{s \in S}$, we define the family $\{\tilde{u}(\cdot, s)\}_{s \in S}$ by $\tilde{u}(x, s) = [\lambda(s)/\lambda^*(s)]u(x, s)$. The ex ante utility of the decision rule ϕ under H , when the agent's utility is \tilde{u} , may be written as

$$\tilde{\mathcal{U}}(\phi, H, \lambda^*) = \sum_{s \in S} \lambda^*(s) \int \tilde{u}(\phi(z), s) dH(z|s).$$

Clearly,

$$\mathcal{U}(\phi, H, \lambda) \equiv \sum_{s \in S} \lambda(s) \int u(\phi(z), s) dH(z|s) = \tilde{\mathcal{U}}(\phi, H, \lambda^*).$$

From this, we conclude that

$$\mathcal{V}(H, \lambda, u) = \mathcal{V}(H, \lambda^*, \tilde{u}). \quad (18)$$

Crucially, the fact that $\{u(\cdot, s)\}_{s \in S}$ is an SCP family, guarantees that $\{\tilde{u}(\cdot, s)\}_{s \in S}$ is

also an SCP family. By assumption, $\mathcal{V}(H, \lambda^*, \tilde{u}) \geq \mathcal{V}(G, \lambda^*, \tilde{u})$. Applying (18) to both sides of this inequality, we obtain $\mathcal{V}(H, \lambda, u) \geq \mathcal{V}(G, \lambda, u)$. QED

Loosely speaking, this result says that if we wish to have ex ante utility comparability for *any* SCP family (or, even more strongly, any IDO family), then fixing the prior does not lead to a weaker criterion of informativeness. A weaker criterion can only be obtained if we fix the prior *and* require ex ante utility comparability for a smaller class of utility families.¹⁷

To construct a converse to Theorem 3, we assume that there are two states and two actions and that the actions are *non-ordered with respect to u* in the sense that x_1 is the better action in state s_1 and x_2 the better action in s_2 , i.e., $u(x_1, s_1) > u(x_2, s_1)$ and $u(x_1, s_2) < u(x_2, s_2)$. This condition guarantees that information on the state is potentially useful; if it does not hold, the decision problem is clearly trivial since either x_1 or x_2 will be unambiguously superior to the other action. Note also that the family $\{u(\cdot, s_1), u(\cdot, s_2)\}$ is an IDO family. We have the following result.

PROPOSITION 8: *Suppose that $S = \{s_1, s_2\}$, $X = \{x_1, x_2\}$, and that the actions are non-ordered with respect to u . If H is MLR-ordered and not more accurate than G , then there is a prior $\bar{\lambda}$ on S such that $\mathcal{V}(H, \bar{\lambda}, u) < \mathcal{V}(G, \bar{\lambda}, u)$.*

Proof: Since H is not more accurate than G , there is \bar{z} and \bar{t} such that

$$G(\bar{z}|s_1) = H(\bar{t}|s_1) \quad \text{and} \quad G(\bar{z}|s_2) < H(\bar{t}|s_2). \quad (19)$$

Given any prior λ , and with the information structure H , we may work out the posterior distribution and the posterior expected utility of any action after receipt of a signal. We claim that there is a prior $\bar{\lambda}$ such that, action x_1 maximizes the agent's posterior expected utility after he receives the signal $z < \bar{t}$ (under H), and action x_2 maximizes the agent's posterior expected utility after he receives the signal $z \geq \bar{t}$. This result follows from the assumption that H is MLR-ordered and is proved in Appendix B.

¹⁷This possibility is explored in Athey and Levin (2001).

Therefore, the decision rule ϕ such that $\phi(z) = x_1$ for $z < \bar{t}$ and $\phi(z) = x_2$ for $z \geq \bar{t}$ maximizes the agent's ex ante utility, i.e.

$$\begin{aligned}\mathcal{V}(H, \bar{\lambda}, u) &= \mathcal{U}(\phi, H, \bar{\lambda}) \\ &= \bar{\lambda}(s_1) \{ u(x_1|s_1)H(\bar{t}|s_1) + u(x_2|s_1)[1 - H(\bar{t}|s_1)] \} \\ &\quad + \bar{\lambda}(s_2) \{ u(x_1|s_2)H(\bar{t}|s_2) + u(x_2|s_2)[1 - H(\bar{t}|s_2)] \}.\end{aligned}$$

Now consider the decision rule ψ under G given by $\psi(z) = x_1$ for $z < \bar{z}$ and $\psi(z) = x_2$ for $z \geq \bar{z}$. We have

$$\begin{aligned}\mathcal{U}(\psi, G, \bar{\lambda}) &= \bar{\lambda}(s_1) \{ u(x_1|s_1)G(\bar{z}|s_1) + u(x_2|s_1)[1 - G(\bar{z}|s_1)] \} \\ &\quad + \bar{\lambda}(s_2) \{ u(x_1|s_2)G(\bar{z}|s_2) + u(x_2|s_2)[1 - G(\bar{z}|s_2)] \}.\end{aligned}$$

Comparing the expressions for $\mathcal{U}(\psi, G, \bar{\lambda})$ and $\mathcal{U}(\phi, H, \bar{\lambda})$, bearing in mind (19), and the fact that x_2 is the optimal action in state s_2 , we see that

$$\mathcal{U}(\psi, G, \bar{\lambda}) > \mathcal{U}(\phi, H, \bar{\lambda}) = \mathcal{V}(H, \bar{\lambda}, u).$$

Therefore, $\mathcal{V}(G, \bar{\lambda}, u) > \mathcal{V}(H, \bar{\lambda}, u)$.

QED

6. STATISTICAL DECISION THEORY

Much of what is known and used in economics on comparative information have their origins in statistical decision theory, so it is appropriate for us to re-present and further develop the results of the last section in that context.¹⁸ The main result in this section is a *complete class theorem*, which says that, in a certain sense and under certain conditions (which we will make precise), the statistician need only employ increasing decision rules.

We assume that the statistician conducts an experiment in which she observes the realization of a random variable (i.e., the outcome of the experiment) that takes values

¹⁸For an introduction to statistical decision theory, see Blackwell and Girshik (1954) or Berger (1985).

in a compact interval Z in R . The distribution of this random variable depends on the state s ; we denote this distribution by $H(\cdot|s)$ and assume that it admits a density function and has Z as its support. The state s is chosen by Nature from a set S , also contained in R . As in the previous section, S may either be a compact interval or a finite set of points. Unlike the last section, we now adopt the perspective of a classical rather than Bayesian statistician, so we do not assume at the outset the existence of a (prior) probability distribution on S .

After observing the experiment's outcome, the statistician takes a decision from a compact set X contained in R ; formally, a decision function (or rule) is a measurable map ϕ from Z to X . Associated to each action and state is a loss; the *loss function* L maps $X \times S$ to R . We assume that L is continuous in the action x . Let D be the set of all decision functions. This experiment, which we shall call H , has the *risk function* $\mathcal{R}_H : D \times S \rightarrow R$, given by

$$\mathcal{R}_H(\phi, s) = \int_{z \in Z} L(\phi(z), s) dH(z|s).$$

We refer to $\mathcal{R}_H(\phi, s)$ as *the risk of ϕ in states s* . For a classical statistician, the risk function \mathcal{R}_H is the central concept with which decisions and experiments are compared. Note that our assumptions guarantee that this function is well defined.

Comparison of Experiments

Suppose the statistician may conduct another experiment G that also has outcomes in Z . At each state s , the distribution $G(\cdot|s)$ admits a density function and has Z as its support. We denote G 's risk function by \mathcal{R}_G . Our first result of this section is a re-statement of Proposition 6 in the context of statistical decisions.

PROPOSITION 9: *Suppose $\{-L(\cdot, s)\}_{s \in S}$ is an IDO family and H is more accurate than G . Then for any increasing decision rule $\psi : Z \rightarrow X$ under G , there is an increasing decision rule $\phi : Z \rightarrow X$ under H such that*

$$\mathcal{R}_H(\phi, s) \leq \mathcal{R}_G(\psi, s) \text{ at each } s \in S. \tag{20}$$

This proposition considers the case when $\{-L(\cdot, s)\}_{s \in S}$ is an IDO family. It says that for any increasing decision rule employed in experiment G , there is a decision rule under H which gives lower risk at every possible state. For this result to be interesting, we must explain why allowing the statistician to employ other decision rules under G (besides increasing rules) is of no use to her. If we can show this, then experiment H is indeed superior to G as measured by risk.

We have already implicitly given (in the previous section) the justification for the focus on increasing decision rules in the case of a Bayesian statistician. The Bayesian will have a prior on S , given by the probability distribution λ , and is interested in decision rules ψ that minimize the expected or *Bayes risk*, which is $\int_{s \in S} \mathcal{R}_G(\psi, s) d\lambda$. Note that

$$\int_{s \in S} \mathcal{R}_G(\psi, s) d\lambda = \int_{z \in Z} \left[\int_{s \in S} L(\psi(z), s) d\lambda_G^z \right] dM_{G, \lambda}$$

where λ_G^z is the posterior distribution of s given z and $M_{H, \lambda}$ is the marginal distribution of z . Suppose $\{G(\cdot | s)\}_{s \in S}$ is MLR-ordered. Then $\{\lambda_G^z\}_{z \in Z}$ is MLR-ordered (in z). Since $\{-L(\cdot, s)\}_{s \in S}$ is an IDO family, it follows from Theorem 2 that the optimal decision rule ψ can indeed be chosen to be increasing.¹⁹ Thus, when presented with the alternative experiments H and G obeying the conditions of Proposition 9, the Bayesian statistician will certainly prefer H since it follows immediately from (20) that H gives a lower Bayes risk.

For the statistician who chooses not to use Bayes risk as her criterion, a different justification must be given for the focus on increasing decision rules.

The completeness of increasing decision rules

We confine our attention to experiment G . The decision rule ψ is said to be *at least as good as* another decision $\tilde{\psi}$, if it has lower risk in all states, i.e., $\mathcal{R}_G(\psi, s) \leq \mathcal{R}_G(\tilde{\psi}, s)$ for all s in S . A subset D' of decision rules forms an *essentially complete class* if for

¹⁹Note that our argument here is completely analogous to the one used to establish Theorem 3.

any decision rule $\tilde{\psi}$, there is a rule ψ in D' that is at least as good as $\tilde{\psi}$. Results which identify some subset of decision rules as an essentially complete class are called *complete class theorems*.

It is useful to identify such a class of decision rules because, while statisticians may differ on the criterion they adopt in choosing amongst decision rules, it is typically the case that a rule satisfying their preferred criterion can be found in an essentially complete class. This is clear for the Bayesian statistician: if $\tilde{\psi}$ minimizes Bayes risk, then a rule ψ in the complete class that is at least as good as $\tilde{\psi}$ will also minimize Bayes risk. The non-Bayesian statistician will choose a rule using a different criterion; the two most commonly used are the *minimax* and *minimax regret* criteria. The minimax criterion evaluates a decision rule according to the largest loss the rule can incur; formally, a rule ψ satisfies this criterion if it solves $\min_{\psi \in D} \{\max_{s \in S} \mathcal{R}_G(\psi, s)\}$. The em regret of a decision rule ψ , which we denote by $r(\psi)$, is defined as $\max_{s \in S} [\mathcal{R}_G(\psi, s) - \min_{\psi' \in D} \mathcal{R}_G(\psi', s)]$. A rule $\hat{\psi}$ satisfies the minimax regret criterion if it solves $\min_{\psi \in D} r(\psi)$.²⁰

The following complete class theorem is the main result of this section.

THEOREM 4: *Suppose $\{-L(\cdot, s)\}_{s \in S}$ is an IDO family and G is MLR-ordered. Then the increasing decision rules form an essentially complete class.*

This result provides a justification for restricting our attention to increasing decision rules that will satisfy both the Bayesian and the classical statistician. Put another way, the statistician searching for an optimal rule need only search amongst increasing decision rules, whether she is using the Bayesian, minimax, or minimax regret criterion.²¹

²⁰In the context of statistical decisions, the minimax criterion was first studied by Wald; the minimax regret criterion is due to Savage. Discussion and motivation for the minimax and minimax regret criteria can be found in Blackwell and Girshik (1954), Berger (1985), and Manski (2005). For some recent applications of the minimax regret criterion, see Manski (2004, 2005) and Manski and Tetenov (2007); a closely related criterion is employed in Chamberlain (2000).

²¹It is worth pointing out another complete class theorem. Some readers may wonder why, in our

Theorem 4 generalizes the complete class theorem of Karlin and Rubin (1956), which in turn generalizes Blackwell and Girshik (1954, Theorem 7.4.3). Karlin and Rubin (1956) establishes the essential completeness of increasing decision rules under the assumption that $\{-L(\cdot, s)\}_{s \in S}$ is a QCIP family, which is a special case of our assumption that $\{-L(\cdot, s)\}_{s \in S}$ forms an IDO family. Note that Theorem 4 is not known even for the case where $\{-L(\cdot, s)\}_{s \in S}$ forms an SCP family. Combining Theorem 4 and Proposition 9 tells us that when an experiment H is more accurate than G , then H is capable of achieving lower risks at all states. So we have discovered something about the classical statistician that we already know about her Bayesian counterpart: she too regards H as superior to G .

COROLLARY 1: *Suppose $\{-L(\cdot, s)\}_{s \in S}$ is an IDO family, H is more accurate than G , and G is MLR-ordered. Then for any decision rule $\psi : Z \rightarrow X$ under G , there is an increasing decision rule $\phi : Z \rightarrow X$ under H such that*

$$\mathcal{R}_H(\phi, s) \leq \mathcal{R}_G(\psi, s) \text{ at each } s \in S. \quad (21)$$

Proof: If ψ is not increasing, then by Theorem 4, there is an increasing rule $\bar{\psi}$ that is at least as good as ψ . Proposition 8 in turn guarantees that there is a decision rule under H that is at least as good as $\bar{\psi}$. QED

Corollary 1 is a generalization of Lehmann (1988) which establishes a version of this result in the case where $\{-L(\cdot, s)\}_{s \in S}$ is a QCIP family. Note that Corollary 1 is more general than Theorem 3 because it establishes the superiority of H over G under a more stringent criterion (see (21)) than Bayes risk. This result is not known even for the case where $\{-L(\cdot, s)\}_{s \in S}$ forms an SCP family.

definition of a decision rule, we did not allow the statistician to mix actions at any given signal. The answer is that when the signal space is atomless (as we have assumed), allowing her to do so will not make a difference, since the set of decision rules involving only pure actions form an essentially complete class (see Blackwell (1951)).

Proof of Theorem 4: The idea of the proof is to show that the statistician who uses a strategy that is not increasing is in some sense debasing the information made available to her by G .

Having assumed that Z is a compact interval, we can assume, without further loss of generality, that $Z = [0, 1]$. Suppose that ψ is a decision rule (not necessarily increasing) under G . Here we confine ourselves to the case where ψ takes only finitely many values, an assumption which certainly holds if the set of actions X is finite. The case where ψ has an infinite range is covered in Appendix B.

Suppose that the actions taken under ψ are exactly x_1, x_2, \dots, x_n (arranged in increasing order). We construct a new experiment \bar{G} along the following lines. For each s , $\bar{G}(0|s) = 0$ and $\bar{G}(k/n|s) = \Pr_G[\psi(z) \leq x_k|s]$ for $k = 1, 2, \dots, n$, where the right side of the second equation refers to the probability of $\{z \in Z : \psi(z) \leq x_k\}$ under the distribution $G(\cdot|s)$. We define $t_k(s)$ as the unique element in Z that obeys $G(t_k(s)|s) = \bar{G}(k/n|s)$. (Note that $t_0(s) = 0$ for all s .) Any z in $((k-1)/n, k/n)$ may be written as $z = \theta[(k-1)/n] + (1-\theta)[k/n]$ for some θ in $(0, 1)$; we define

$$\bar{G}(z|s) = G(\theta t_{k-1}(s) + (1-\theta)t_k(s)|s). \quad (22)$$

This completely specifies the experiment \bar{G} .

Define a new decision rule $\bar{\psi}$ by $\bar{\psi}(z) = x_1$ for z in $[0, 1/n]$; for $k \geq 2$, we have $\bar{\psi}(z) = x_k$ for z in $((k-1)/n, k/n]$. This is an increasing decision rule under \bar{G} . It is also clear from our construction of \bar{G} and $\bar{\psi}$ that, at each state s , the distribution of utility induced by \bar{G} and $\bar{\psi}$ equals the distribution of utility induced by G and ψ .

We claim that G is more accurate than \bar{G} . Provided this is true, Proposition 6 says that there is an increasing decision rule ϕ under G that is at least as good as $\bar{\psi}$ under \bar{G} , i.e., at each s , the distribution of utility induced by G and ϕ first order stochastically dominates that induced by \bar{G} and $\bar{\psi}$. Since the latter coincides with the distribution of utility induced by G and ψ , the proof is complete.

That G is more accurate than \bar{G} follows from the assumption that G is MLR-

ordered. We prove this in Appendix B.

QED

It is clear from our proof of Theorem 4 that we can in fact give a sharper statement of that result; we do so below.

THEOREM 4*: *Suppose $\{-L(\cdot, s)\}_{s \in S}$ is an IDO family and G is MLR-ordered. Then for any decision rule $\psi : Z \rightarrow X$, there is increasing decision rule $\phi : Z \rightarrow X$ such that, at each s , the distribution of utility induced by G and ϕ first order stochastically dominates the distribution of utility induced by G and ψ .*

Example 6 continued. Recall that we assumed in this application that the manager's decision rule under G , which is ψ , is increasing in the signal. Provided G is MLR-ordered, Theorem 4* provides a justification for this rule.

Let $\tilde{\psi}$ be a (not necessarily increasing) decision rule. Theorem 4* tells us that for some investor k , there is an increasing decision rule $\psi : Z \rightarrow X$ such that, at each s , the distribution of u_k induced by G and ψ first order stochastically dominates the distribution of u_k induced by G and $\tilde{\psi}$. This implies that

$$\int_{z \in Z} u_k(\phi(z), s) dG(z|s) \geq \int_{z \in Z} u_k(\tilde{\psi}(z), s) dG(z|s). \quad (23)$$

Aggregating this inequality across states, we obtain $\mathcal{U}_k(\psi, G, \lambda_k) \geq \mathcal{U}_k(\tilde{\psi}, G, \lambda_k)$, i.e., the increasing rule gives investor k a higher ex ante utility.

However, we can say more because, for any other investor i , $u_i(\cdot, s)$ is just an increasing transformation of $u_k(\cdot, s)$, i.e., there is a strictly increasing function f such that $u_i = f \circ u_k$. Appealing to Theorem 4* again, we see that (23) is true if u_k is replaced with u_i . Aggregating this inequality across states gives us $\mathcal{U}_i(\psi, G, \lambda_i) \geq \mathcal{U}_i(\tilde{\psi}, G, \lambda_i)$. In short, we have shown the following: any decision rule admits an increasing decision rule that (weakly) raises the ex ante utility of *every* investor. This justifies our assumption that the manager uses an increasing decision rule.

Our final application generalizes an example found in Manski (2005, Proposition 3.1) on monotone treatment rules.

Example 7. Suppose that there are two ways of treating patients with a particular medical condition. Treatment A is the status quo; it is known that a patient who receives this treatment will recover with probability \bar{p}^A . Treatment B is a new treatment whose effectiveness is unknown. The probability of recovery with this treatment, p^B , corresponds to the unknown state of the world and takes values in some set P . We assume that the planner receives a signal z of p^B that is MLR-ordered with respect to p^B . (Manski (2005) considers the case of N subjects who are randomly selected to receive Treatment B; with z being the the number who are cured. Clearly, the distribution of z is binomial; it is also not hard to check that it is MLR-ordered with respect to p^B .)

Normalizing the utility of a cure at 1 and that of no cure at 0, the planner's expected utility when a member of the population receives treatment A is \bar{p}^A . Similarly, the expected utility of treatment B is p^B . Therefore, the planner's utility if she subjects fraction x of the population to B (and the rest to A) is

$$u(x, p^B) = (1 - x)\bar{p}^A + x p^B. \quad (24)$$

The planner's decision (treatment) rule maps z to the proportion x of the (patient) population who will receive treatment B. As pointed out in Manski (2005), $\{u(x, \cdot)\}_{p^B \in P}$ is a QCIP family, and so Karlin and Rubin (1956) guarantee that decision rules where x increases with z form an essentially complete class.

Suppose now that the planner has a different payoff function, that takes into account the cost of the treatment. We denote the cost of having fraction x treated with B and the rest with A by $C(x)$. Then the payoff function is

$$u(x, p^B) = (1 - x)\bar{p}^A + x p^B - C(x).$$

If the cost of treatments A and B are both linear, or more generally if C is convex, then one can check that $\{u(x, \cdot)\}_{p^B \in P}$ will still be a QCIP family. We can then appeal to Karlin and Rubin (1956) to obtain the essential completeness of the increasing

decision rules. But there is no particular reason to believe that C is convex; indeed C will never be convex if the presence of scale economies leads to the total cost of having both treatments in use being more expensive than subjecting the entire population to one treatment or the other. (Formally, there is $x^* \in (0, 1)$ such that $C(0) < C(x^*) > C(1)$.) However, u is supermodular in (x, p^B) whatever the shape of C , so $\{u(x, \cdot)\}_{p^B \in P}$ is certainly an IDO family; Theorem 4 tells us that the planner may confine herself to increasing decision rules since they form an essentially complete class.²²

APPENDIX A

Our objective in this section is to prove Theorem 2 in the case where S is a finite set. Suppose that $S = \{s_1, s_2, \dots, s_N\}$ and the agent's objective function is

$$F(x, \lambda) = \sum_{i=1}^N f(x, s_i) \lambda(s_i)$$

where $\lambda(s_i)$ is the probability of state s_i . We assume that $\lambda(s_i) > 0$ for all s_i . Let γ be another distribution with support S . We say that γ is an MLR-shift of λ if $\gamma(s_i)/\lambda(s_i)$ is increasing in i .

Proof of Theorem 2 for the case of finite S : Suppose $F(x'', \lambda) \geq F(x, \lambda)$ for x in $[x', x'']$. We denote $(f(x'', s_i) - f(x', s_i))\lambda(s_i)$ by a_i and define $A_k = \sum_{i=k}^N a_i$. By assumption, $A_1 = F(x'', \lambda) - F(x', \lambda) \geq 0$; we claim that $A_k \geq 0$ for any k (this claim is analogous to (7) in the proof of Theorem 2).

²²Manski and Tetenov (2007) prove a complete class theorem with a different variation on the payoff function (24). For the payoff function (24), Manski (2005) in fact showed a sharper result: the planner will choose a rule where the *whole* population is subject to B (A) if the number of treatment successes in the sample goes above (below) a particular threshold. The modification of the payoff function in Manski and Tetenov (2007) is motivated in part by the desire to obtain fractional treatment rules, in which, for a non-negligible set of sample outcomes, *both* treatments will be in use. In our variation on (24), it is clear that fractional treatment is plausible if large values of x involve very high costs.

Suppose instead that there is $M \geq 2$ such that

$$A_M = \sum_{i=M}^N (f(x'', s_i) - f(x', s_i)) \lambda(s_i) < 0.$$

As in the proof of Theorem 2 in the main part of the paper, we choose \bar{x} that maximizes $f(\cdot, s_M)$ in $[x', x'']$. By the IDO property and Lemma 2, we have

$$f(\bar{x}, s_i) - f(x'', s_i) \geq 0 \text{ for } i \leq M \text{ and} \quad (25)$$

$$f(\bar{x}, s_i) - f(x', s_i) \geq 0 \text{ for } i \geq M. \quad (26)$$

(These inequalities (25) and (26) are analogous to (9) and (10) respectively.) Following the argument used in Theorem 2, these inequalities lead to

$$A_1 = \sum_{i=1}^N (f(x'', s_i) - f(x', s_i)) \lambda(s_i) < 0, \quad (27)$$

which is a contradiction. Therefore $A_M \geq 0$.

Denoting $\gamma(s_i)/\lambda(s_i)$ by b_i , we have

$$F(x'', \gamma) - F(x', \gamma) = \sum_{i=1}^N a_i b_i.$$

It is not hard to check that²³

$$A_1 b_1 + \sum_{i=2}^N A_i (b_i - b_{i-1}) = \sum_{i=1}^N a_i b_i. \quad (28)$$

Since γ is an MLR shift of λ , $b_i - b_{i-1} \geq 0$ for all i and so $\sum_{i=2}^N A_i (b_i - b_{i-1}) \geq 0$.

Thus (28) guarantees that $\sum_{i=1}^N a_i b_i \geq A_1 b_1$; in other words,

$$F(x'', \gamma) - F(x', \gamma) \geq \frac{\gamma(s_1)}{\lambda(s_1)} [F(x'', \lambda) - F(x', \lambda)]. \quad (29)$$

It follows that the left hand side is nonnegative (positive) if the right side is nonnegative (positive), as required by the IDO property. QED

²³This is just a discrete version of integration by parts.

APPENDIX B

Proof of Lemma 3 continued: We first show that ϕ (as constructed in the proof in Section 5) is an increasing rule. We wish to compare $\phi(t')$ against $\phi(t)$ where $t' > t$. Note that the construction of $\phi(t)$ first involves partitioning S into subsets S_1, S_2, \dots, S_M obeying properties (i) to (v). In particular, (v) says that for any s in S_m , $\psi(\tau(t, s))$ takes the same value, which we denote by ψ_m . To obtain $\phi(t')$, we first partition S into disjoint sets S'_1, S'_2, \dots, S'_L , where L is odd, with the partition satisfying properties (i) to (v). The important thing to note is that, for any s , we have

$$\psi(\tau(t', s)) \geq \psi(\tau(t, s)). \quad (30)$$

This is clear: both ψ and $\tau(\cdot, s)$ are increasing functions and $t' > t$. We denote $\psi(\tau(t', s))$ for s in S'_k by ψ'_k . Any s in S belongs to some S_m and some S'_k , in which case (30) may be re-written as

$$\psi'_k \geq \psi_m. \quad (31)$$

The construction of $\phi(t')$ involves the construction of $\hat{\phi}'_2, \hat{\phi}'_4$, etc. The action $\hat{\phi}'_2$ is the largest action maximizing $u(\cdot, s^*)$ in the interval $[\psi'_2, \psi'_1]$. Comparing this with $\hat{\phi}_2$, which is the largest action maximizing $u(\cdot, s^*)$ in the interval $[\psi_2, \psi_1]$, we know that $\hat{\phi}'_2 \geq \hat{\phi}_2$ since (following from (31)) $\psi'_2 \geq \psi_2$ and $\psi'_1 \geq \psi_1$.

By definition, $\hat{\phi}'_4$ is the largest action maximizing $u(\cdot, s'_3)$ in $[\psi'_4, \hat{\phi}'_2]$, where s'_3 refers to the unique element in S'_3 . Let \bar{m} be the largest odd number such that $s_{\bar{m}} \leq s'_3$. (Recall that $s_{\bar{m}}$ is the unique element in $S_{\bar{m}}$.) By definition, $\hat{\phi}_{\bar{m}+1}$ is the largest element maximizing $u(\cdot, s_{\bar{m}})$ in $[\psi_{\bar{m}+1}, \hat{\phi}_{\bar{m}-1}]$. We claim that $\hat{\phi}'_4 \geq \hat{\phi}_{\bar{m}+1}$. This is an application of Theorem 1. It follows from the following: (i) $s'_3 \geq s_{\bar{m}}$, so $u(\cdot, s'_3) \succeq u(\cdot, s_{\bar{m}})$; (ii) the manner in which \bar{m} is defined, along with (31), guarantees that $\psi'_4 \geq \psi_{\bar{m}+1}$; and (iii) we know (from the previous paragraph) that $\hat{\phi}'_2 \geq \hat{\phi}_2 \geq \hat{\phi}_{\bar{m}-1}$.

So we obtain $\hat{\phi}'_4 \geq \hat{\phi}_{\bar{m}+1} \geq \phi(t)$. Repeating the argument finitely many times (on $\hat{\phi}'_6$ and so on), we obtain $\phi(t') \geq \phi(t)$. This completes the proof of Lemma 3 in the

special case where ψ takes finitely many values.

Extension to the case where the range of ψ is infinite.

The strategy is to approximate ψ with a sequence of simpler decision rules. Let $\{A_n\}_{n \geq 1}$ be a sequence of finite subsets of X such that $A_n \subset A_{n+1}$ and $\cup_{n \geq 1} A_n$ is dense in X . (This sequence exists because X is compact.) The function $\psi_n : Z \rightarrow X$ is defined as follows: $\psi_n(z)$ is the largest element in A_n that is less than or equal to $\psi(z)$. The sequence of decision rules ψ_n has the following properties:

- (i) ψ_n is increasing in z ;
- (ii) $\psi_{n+1}(z) \geq \psi_n(z)$ for all z ;
- (iii) the range of ψ_n is finite; and
- (iv) the increasing sequence ψ_n converges to ψ pointwise.

Since ψ_n takes only finitely many values, we know there is an increasing decision rule ϕ_n (as defined in the proof of Lemma 3 in Section 5) such that for all (z, s) ,

$$u(\phi_n(T(z, s)), s) \geq u(\psi_n(z), s). \quad (32)$$

We claim that ϕ_n is also an increasing sequence. This follows from the fact that, for all (t, s) ,

$$\psi_{n+1}(\tau(t, s)) \geq \psi_n(\tau(t, s)). \quad (33)$$

This inequality plays the same role as (30); the latter was used to show that $\phi(t') \geq \phi(t)$. Mimicking the argument there, (33) tells us that $\phi_{n+1}(t) \geq \phi_n(t)$.

Since ϕ_n is an increasing sequence and X is compact, it has a limit, which we denote as ϕ . Since, for each n , ϕ_n is an increasing decision rule, ϕ is also an increasing decision rule. For each n , (32) holds; taking limits, and using the continuity of u with respect to x , we obtain $u(\phi(T(z, s)), s) \geq u(\psi(z), s)$. QED

Proof of Proposition 7 continued: It remains for us to show how $\bar{\lambda}$ is constructed. We denote the density function of $H(\cdot|s)$ by $h(\cdot|s)$. It is clear that since the actions are non-ordered, we may choose $\bar{\lambda}(s_1)$ and $\bar{\lambda}(s_2)$ such that

$$\bar{\lambda}(s_1)h(\bar{t}|s_1) [u(x_1, s_1) - u(x_2, s_1)] = \bar{\lambda}(s_2)h(\bar{t}|s_2) [u(x_2|s_2) - u(x_1, s_2)]. \quad (34)$$

Re-arranging this equation, we obtain

$$\bar{\lambda}(s_1)h(\bar{t}|s_1)u(x_1|s_1)+\bar{\lambda}(s_2)h(\bar{t}|s_2)u(x_1|s_2) = \bar{\lambda}(s_1)h(\bar{t}|s_1)u(x_2|s_1)+\bar{\lambda}(s_2)h(\bar{t}|s_2)u(x_2|s_2).$$

Therefore, given the prior $\bar{\lambda}$, the posterior distribution after observing \bar{t} is such that the agent is indifferent between actions x_1 and x_2 .

Suppose the agent receives the signal $z < \bar{t}$. Since H is MLR-ordered, we have

$$\frac{h(z|s_1)}{h(\bar{t}|s_1)} \geq \frac{h(z|s_2)}{h(\bar{t}|s_2)}.$$

This fact, together with (34) guarantee that

$$\bar{\lambda}(s_1)h(z|s_1) [u(x_1, s_1) - u(x_2, s_1)] \geq \bar{\lambda}(s_2)h(z|s_2) [u(x_2|s_2) - u(x_1, s_2)].$$

Re-arranging this equation, we obtain

$$\begin{aligned} \bar{\lambda}(s_1)h(z|s_1)u(x_1|s_1) + \bar{\lambda}(s_2)h(z|s_2)u(x_1|s_2) &\geq \\ \bar{\lambda}(s_1)h(z|s_1)u(x_2|s_1) + \bar{\lambda}(s_2)h(z|s_2)u(x_2|s_2). & \end{aligned}$$

So, after observing $z < \bar{t}$, the (posterior) expected utility of action x_1 is greater than that of x_2 . In a similar way, we can show that x_2 is the optimal action after observing a signal $z \geq \bar{t}$. QED

Proof of Theorem 4 continued: We denote the density function associated to the distribution $G(\cdot|s)$ by $g(\cdot|s)$. The probability of $Z_k = \{z \in Z : \psi(z) \leq x_k\}$ is given by $\int \mathbf{1}_{Z_k}(z)g(z|s)dz$, where $\mathbf{1}_{Z_k}$ is the indicator function of Z_k . By the definition of \bar{G} , we have

$$\bar{G}(k/n|s) = \int \mathbf{1}_{Z_k}(z)g(z|s)dz.$$

Recall that $t_k(s)$ is defined as the unique element that obeys $G(t_k(s)|s) = \bar{G}(k/n|s)$; equivalently,

$$\bar{G}(k/n|s) - G(t_k(s)|s) = \int [\mathbf{1}_{Z_k}(z) - \mathbf{1}_{[0, t_k(s)]}(z)] g(z|s)dz = 0. \quad (35)$$

The function W given by $W(z) = \mathbf{1}_{Z_k}(z) - \mathbf{1}_{[0, t_k(s)]}(z)$ has the following single-crossing type condition: $z > t_k(s)$, we have $W(z) \geq 0$ and for $z \leq t_k(s)$, we have $W(z) \leq 0$.²⁴ Let $s' > s$; since $G(\cdot|s)$ is MLR-ordered, $g(z|s')/g(z|s)$ is an increasing function of z . By a standard result (see, for example, Athey (2002, Lemma 5)), we have

$$\bar{G}(k/n|s') - G(t_k(s)|s') = \int [\mathbf{1}_{Z_k}(z) - \mathbf{1}_{[0, t_k(s)]}(z)] g(z|s') dz \geq 0. \quad (36)$$

This implies that $t_k(s') \geq t_k(s)$.

To show that G is more accurate than \bar{G} , we require $T(z, s)$ to be increasing in s , where T is defined by $G(T(z, s)|s) = \bar{G}(z|s)$. For $z = k/n$, $T(z, s) = t_k(s)$, which we have shown is increasing in s . For z in the interval $((k-1)/n, k/n)$, recall (see (22)) that $\bar{G}(z)$ was defined such that $T(z, s) = \theta t_{k-1}(s) + (1-\theta)t_k(s)$. Since both t_{k-1} and t_k are increasing in s , $T(z, s)$ is also increasing in s . This complete the proof in the case where ψ takes only finitely many values.

Extension to the case where ψ has an infinite range

We construct an alternative experiment \bar{G} and increasing decision rule $\bar{\psi}$ with the following two properties:

(P1) at each state s , the distribution of losses induced by \bar{G} and $\bar{\psi}$ equals that induced by G and ψ ;

(P2) G is more accurate than \bar{G} .

An application of Proposition 6 then guarantees that there is an increasing decision rule under G that is at least as good as ψ . Thus our proof is essentially the same as the one we gave for the finite case in Section 6, except that construction of \bar{G} and $\bar{\psi}$ is somewhat more complicated.

Since X is compact, there is a smallest compact interval M containing X . At a given state s , we denote the distribution on M induced by G and ψ by $F(\cdot|s)$, i.e., for any x in M , we have $F(x|s) = \Pr_G[\psi(z) \leq x|s]$. There are two noteworthy features

²⁴This property is related to but not the same as the single crossing property we have defined in this paper; Athey (2002) refers to this property as SC1 and the one we use as SC2.

of $\{F(\cdot|s)\}_{s \in S}$:

(i) For a fixed \bar{s} , we may partition M into (disjoint) contour sets, $U_{\bar{s}}(r)$, i.e., $U_{\bar{s}}(r) = \{x \in M : F(x|\bar{s}) = r\}$. It is possible that for some r , $U_{\bar{s}}(r)$ is empty, but if it is nonempty then it has a minimum and the minimum is in X (and not just in M). Crucially, this partition is *common across all states* s . In other words, for any other state s , there is some r' such that $U_s(r') = U_{\bar{s}}(r)$.

(ii) The atoms of $F(\cdot|s)$ also do not vary with s ; i.e., if x is an atom for $F(\cdot|\bar{s})$, then it is an atom for $F(\cdot|s)$ for every other state s .

These two features follow easily from the definition of F , the compactness of X , and the fact that $G(\cdot|s)$ is atomless and has support Z at every state s .

To each element x in M we associate a number $\epsilon(x)$, where $\epsilon(x) > 0$ if and only if x is an atom and $\sum_{x \in X} \epsilon(x) < \infty$. (Note that there are at most countably many atoms, so the infinite summation makes sense.) We define the map $Y : M \rightarrow R$ where $Y(x) = x + \sum_{\{x' \in M : x' \leq x\}} \epsilon(x')$. It is clear that this map is a strictly increasing and hence 1-1 map. Let $Y^* = \cup_{x \in M} [Y(x) - \epsilon(x), Y(x)]$. The difference between Y^* and the range of Y , i.e., set $Y^* \setminus Y(M)$, may be written in the form $\cup_{n=1}^{\infty} I_n$, where $I_n = [Y(a_n) - \epsilon(a_n), Y(a_n))$ and $\{a_n\}_{n \in N}$ is the set of atoms. (Loosely speaking, the ‘gaps’ I_n arise at every atom.)

We define the distribution $\tilde{G}(\cdot|s)$ on Y^* in the following way. For y in $Y(M)$, $\tilde{G}(y|s) = F(Y^{-1}(y)|s)$. For $y = Y(a_n) - \epsilon(a_n)$, define $\tilde{G}(y|s)$ as the limit of $\tilde{G}(y_n|s)$ where y_n is some sequence in $Y(M)$ tending to y from the left; if no such sequence exists (which occurs if and only if there is an atom at the smallest element of X), let $\tilde{G}(y|s) = 0$. (One can easily check that this definition is unambiguous.) It remains for us to define $\tilde{G}(y|s)$ for y in the open interval $(Y(a_n) - \epsilon(a_n), Y(a_n))$. For $y = C(a_n)$ or $y = C(a_n) - \epsilon(a_n)$, define $t(y)$ by $G(t(y)|s) = \tilde{G}(y|s)$. Any element y in $(Y(a_n) - \epsilon(a_n), Y(a_n))$ may be written as $\theta [Y(a_n) - \epsilon(a_n)] + (1 - \theta) [Y(a_n)]$. We define

$$\tilde{G}(y|s) = G(\theta t(C(a_n) - \epsilon(a_n)) + (1 - \theta)t(C(a_n)) |s).$$

We have now completely specified the distribution $\tilde{G}(\cdot|s)$. Note that we have constructed this distribution to be atomless, so for every number r in $[0, 1]$, the set $\{y \in Y^* : \tilde{G}(y|s) = r\}$ is nonempty. Indeed, following from observation (i) above, this set has a smallest element, which we denote by $\hat{y}(r)$. We define $Y^{**} = \{\hat{y}(r) : r \in [0, 1]\}$. Observation (i) also tells us that Y^{**} does not vary with s . We denote the restriction of $\tilde{G}(\cdot|s)$ to Y^{**} by $\bar{G}(\cdot|s)$. Therefore, for any r in $[0, 1]$ and any state s , there is a unique y in Y^{**} such that $\bar{G}(y|s) = r$.

One can check that property (P2) (stated above) holds: G is more accurate than \bar{G} . Formally, the map T defined by $G(T(y, s)|s) = \bar{G}(y|s)$ exists and has the property that $T(y, s)$ is increasing in s ; the proof is substantially the same as that for the finite case. Furthermore, the map $T(\cdot|s)$ has a unique inverse (in Y^{**}). So we have identified precisely the properties of T needed for the application of Proposition 6.

Consider the decision rule $\bar{\psi} : Y^{**} \rightarrow X$ defined as follows: if y is in $Y(M)$, define $\bar{\psi}(y) = Y^{-1}(y)$; if y is in $[Y(a_n) - \epsilon(a_n), Y(a_n))$, define $\bar{\psi}(y) = a_n$. It is not hard to verify that \bar{G} and $\bar{\psi}$ generate the same distribution of losses as G and ψ , as required by (P1). QED

REFERENCES

- AMIR, R. (1996): "Sensitivity Analysis in Multisector Optimal Economic Dynamics" *Journal of Mathematical Economics*, 25, 123-141.
- ARROW, K. AND D. LEVHARI (1969): "Uniqueness of the internal rate of return with variable life of investment," *Economic Journal*, 79(315), 560-566.
- ASHWORTH, S. AND E. BUENO DE MESQUITA (2006): "Monotone Comparative Statics in Models of Politics," *American Journal of Political Science* 50(1), 214-231.
- ATHEY, S (2002): "Monotone Comparative Statics under Uncertainty," *Quarterly Journal of Economics*, 117(1), 187-223.

- ATHEY, S. AND J. LEVIN (2001): "The Value of Information in Monotone Decision Problems," Stanford Working Paper 01-003.
- ATHEY, S., P. MILGROM, AND J. ROBERTS (1998): *Robust Comparative Statics*. (Draft Chapters.) <http://www.stanford.edu/~athey/draftmonograph98.pdf>
- BECKER, R. AND J. BOYD (1997): *Capital Theory, Equilibrium Analysis, and Recursive Utility*. Oxford: Blackwell.
- BERGEMANN, D. and J. VALIMAKI (2002): "Information Acquisition and Efficient Mechanism Design," *Econometrica*, 70, 1007-1033.
- BERGER, J. O. (1985): *Statistical Decision Theory and Bayesian Analysis* New York: Springer-Verlag
- BLACKWELL, D. (1951): "On a Theorem of Lyapunov," *The Annals of Mathematical Statistics*, 22(1), 112-114.
- BLACKWELL, D. (1953): "Equivalent Comparisons of Experiments," *The Annals of Mathematical Statistics*, 24(2), 265-272.
- BLACKWELL, D. AND M. A. GIRSHIK (1954): *Theory of Games and Statistical Decisions*. New York: Dover Publications
- CHAMBERLAIN, G. (2000): "Econometrics and Decision Theory," *Journal of Econometrics*, 95, 255-283.
- GOLLIER, C. (2001): *The Economics of Risk and Time*. Cambridge: MIT Press.
- JEWITT, I. (2006): "Information Order in Decision and Agency Problems," Personal Manuscript.
- KARLIN, S. AND H. RUBIN (1956): "The Theory of Decision Procedures for Distributions with Monotone Likelihood Ratio," *The Annals of Mathematical Statistics*, 27(2), 272-299.

- LEHMANN, E. L. (1988): “Comparing Location Experiments,” *The Annals of Statistics*, 16(2), 521-533.
- LEVIN, J. (2001): “Information and the Market for Lemons,” *Rand Journal of Economics*, 32(4), 657-666.
- MANSKI, C. (2004): “Statistical Treatment Rules for Heterogeneous Populations,” *Econometrica*, 72(4), 1221-1246.
- MANSKI, C. (2005): *Social Choice with Partial Knowledge of Treatment Response*, Princeton: Princeton University Press.
- MANSKI, C. AND A. TETENOV (2007): “Admissible treatment rules for a risk-averse planner with experimental data on an innovation,” *Journal of Statistical Planning and Inference*, 137(6), 1998-2010.
- MILGROM, P. AND J. ROBERTS (1990): “Rationalizability, Learning, and Equilibrium in Games with Strategic Complementarities,” *Econometrica*, 58(6), 1255-1277.
- MILGROM, P. AND C. SHANNON (1994): “Monotone Comparative Statics,” *Econometrica*, 62(1), 157-180.
- ORMISTON, M. B. AND E. SCHLEE (1993): “Comparative Statics under uncertainty for a class of economic agents,” *Journal of Economic Theory*, 61, 412-422.
- PERSICO, N (1996): “Information Acquisition in Affiliated Decision Problems.” Discussion Paper No. 1149, Department of Economics, Northwestern University.
- PERSICO, N (2000): “Information Acquisition in Auctions,” *Econometrica*, 68(1), 135-148.

- QUAH, J. (2007): “The comparative statics of constrained optimization problems,” *Econometrica*, 75(2), 401-431.
- QUAH, J. K.-H. AND B. STRULOVICI (2007): “Comparative Statics with the Interval Dominance Order II,” Incomplete Manuscript. (First version to be posted on Quah’s webpage by early December.)
- TOPKIS, D. M. (1978): “Minimizing a Submodular Function on a lattice,” *Operations Research*, 26, 305-321.
- TOPKIS, D. M. (1998): *Supermodularity and Complementarity*. Princeton: Princeton University Press.
- VIVES, X. (1990): “Nash Equilibrium with Strategic Complementarities,” *Journal of Mathematical Economics*, 19, 305-21.

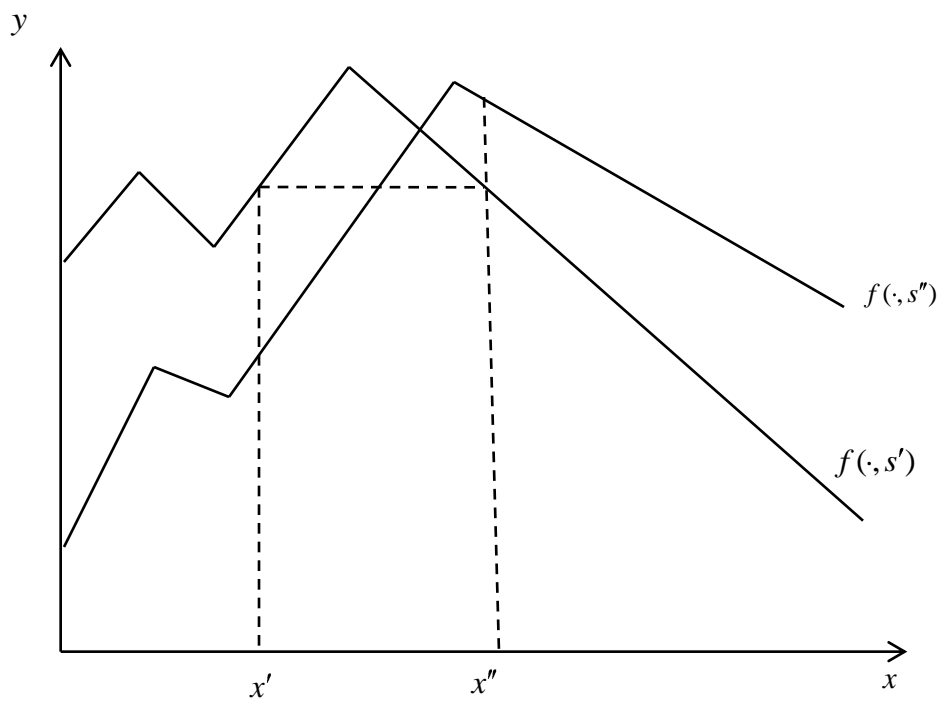


Figure 1

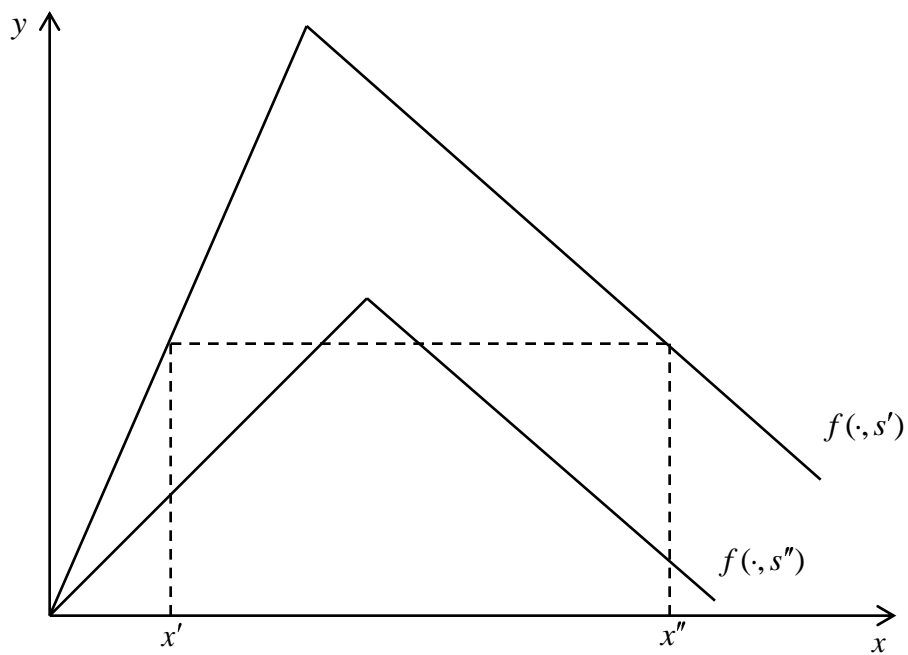


Figure 2