Microeconomic Theory Lecture Notes

Todd Sarver

Duke University, Department of Economics

February 16, 2023

Contents

Preface

These notes grew out of my lectures for the last half-semester of the PhD core microeconomics sequence in the Economics Department at Duke University. The intent of the course (and hence of these notes) is twofold: The first goal is to expose students to some recent topics and tools in microeconomics that are not covered in many traditional textbooks. Second, as the final component of our core PhD sequence, this course aims to synthesize and extend many of the concepts that are covered in earlier parts of the microeconomics sequence.

Broadly speaking, these notes focus on comparisons of risk and information and on how individuals, firms, and other economic agents respond to changes in deterministic or stochastic variables, or to information. The central tools for this type of analysis are referred to as monotone comparative statics, and they can be used to address a wide array of economic questions. For example, under what conditions on a production function will a decrease in the price of labor lead to an increase in the demand for capital? What type of change to the distribution of a risky asset will lead an expected-utility maximizer to increase zir investment in that asset?[1] How do cost changes for one firm affect the equilibrium production decisions of competing firms?

To address such questions, these notes begin with the study of monotone comparative statics in deterministic environments. Even within this relatively simple setting of deterministic single-agent decision problems, the comparative statics results we develop find numerous applications. These concepts are then extended to multi-agent strategic environments, where we obtain equilibrium existence results and comparative statics results for games with strategic complementarities. Next, we expand our analysis to deal with risk. In this segment of the notes, we first revisit several important stochastic orders that may be familiar from prior microeconomics courses: first-order stochastic dominance, the monotone likelihood ratio order, and second-order stochastic dominance. We then explore conditions under which endogenous decision variables respond monotonically to changes in the distribution of exogenous variables according to one of these stochastic orders. The final portion of the notes concerns imperfect information and signals. There, we examine different approaches to modeling information and present Blackwell's classic results on the comparison of experiments.

Some portions of these notes drew inspiration from the excellent treatments offered by other instructors in their teaching materials, including Eddie Dekel, Jonathan Levin, and

[1] I'll often utilize the gender neutral pronouns ze/zir/zirs/zirself.

Muhamet Yildiz. I have also received a great deal of helpful feedback from past students and teaching assistants. A special thanks goes out to Zichang Wang for creating many of the figures for these notes.

Todd Sarver
February 2023

Chapter 1

Monotone Comparative Statics

Contents

References and Assigned Readings

Primary readings:

(1) Milgrom and Shannon (1994) (recommended).
(2) Dekel notes on Monotone Comparative Statics (optional).

Additional references:

(1) Topkis (1998)

1.1 Motivation and Intuition

1.1.1 Motivating Questions

We are often interested in the response of different economic agents to some change in the underlying environment or conditions. We may also be interested in the indirect impact of such changes on the equilibrium behavior of other agents. Monotone comparative statics provides a set of tools that allow us to address these questions and make precise qualitative predictions in many instances.

For a concrete example, consider a firm's production decisions. Monotone comparative statics can be used to answer the following types of questions:

- How does a monopoly production decision respond to a change in production costs?
- What is the response of a firm's demand for one input to a change in the price of that input? Of another input?
- How does a change in the production cost of one firm impact the equilibrium production levels of that and other firms? What role does market structure and the substitutability of these products play in determining this impact?

Broadly speaking, the goal of monotone comparative statics is to give robust qualitative predictions about the interaction between parameters and strategic variables using minimal assumptions.

1.1.2 Implicit Function Theorem Approach

Suppose an agent is maximizing the parameterized objective function $f : X \times T \to \mathbb{R}$, where T is a set of parameters and X is a decision variable. We will begin with the simplest case, where $X \subseteq \mathbb{R}$ and $T \subseteq \mathbb{R}$.

The parameterized **solution set** is defined as follows:

$$X^*(t) = \operatorname*{argmax}_{x \in X} f(x, t).$$

The mapping from t to $X^*(t)$ is a correspondence. A **selection** from $X^*(t)$ is a function

$x^* : T \to \mathbb{R}$ such that $x^*(t) \in X^*(t)$ for all $t \in T$. When there is a unique solution to the maximization problem for every t, we have $X^*(t) = \{x^*(t)\}$.

To make this problem as tractable as possible, we can impose some strong assumptions:

(1) A solution exists for every t. (compactness, continuity)
(2) The objective function f is twice continuously differentiable.
(3) The objective function f is strictly concave in x: $f_{xx} < 0$. (This implies there is a unique solution for every t.)
(4) The solution $x^*(t)$ lies in the interior of X.

Then the solution satisfies the following first-order condition:

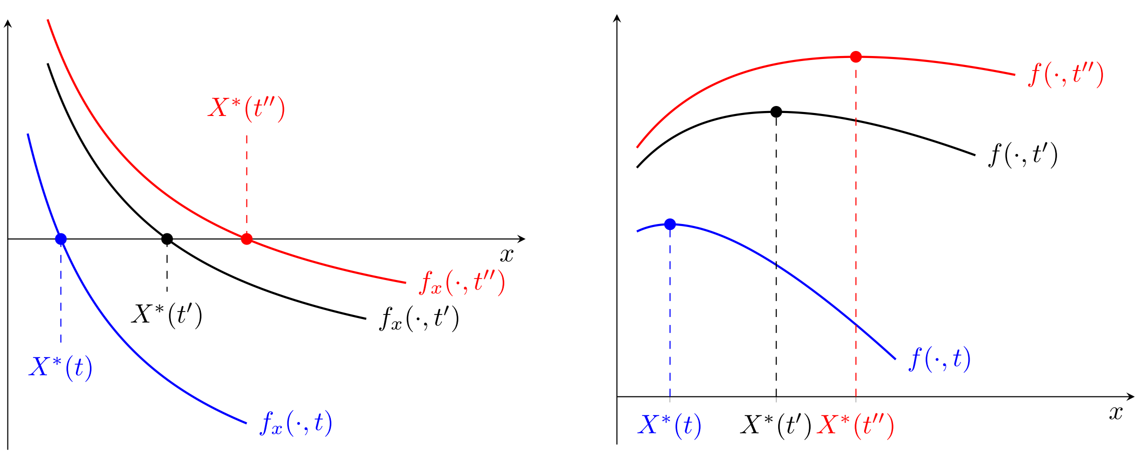
$$f_x(x^*(t), t) = 0. \tag{1.1}$$

In addition, since $f_{xx} < 0$, the implicit function theorem implies that $x^*(t)$ is differentiable in t and[1]

$$x^{*\prime}(t) = -\frac{f_{xt}(x^*(t), t)}{f_{xx}(x^*(t), t)}. \tag{1.2}$$

Thus $x^{*\prime}(t) \geq 0$ if and only if $f_{xt}(x^*(t), t) \geq 0$.

Figure 1.1 illustrates both the partial derivative $f_x(\cdot, t)$ and the objective function $f(\cdot, t)$ for different values of t and shows how the critical points move to the right as t increases, under the assumptions that $f_{xt} > 0$ and $f_{xx} < 0$. The following example illustrates this type of analysis more concretely in the context of a monopoly problem.



(a) Derivative f_x (b) Objective function f

Figure 1.1: Solutions for the parameters $t < t' < t''$ in the case where $f_{xx} < 0$.

[1] Note that the nontrivial part of the implicit function theorem is showing that $x^*(t)$ is differentiable. Once that is established, the exact formula for the derivative follows immediately by applying the chain rule to obtain

$$f_{xx}(x^*(t), t)\, x^{*\prime}(t) + f_{xt}(x^*(t), t) = 0$$

and then rearranging terms.

EXAMPLE 1.1 (MONOPOLY RESPONSE TO CHANGE IN COST). Consider a monopoly in a market where inverse demand is given by $P(x)$. Suppose the firm's total cost function $C(x,t)$ depends on both the quantity produced x and some parameter t. The total profit of the firm is therefore given by

$$f(x,t) = P(x)x - C(x,t).$$

The firm chooses output to maximize profit:

$$X^*(t) = \operatorname*{argmax}_{x \geq 0} f(x,t).$$

Suppose we are interested in how the firm's production decision changes with the parameter t. To determine the relationship, consider the first-order condition for an interior solution:

$$f_x(x,t) = \underbrace{P'(x)x + P(x)}_{MR} - \underbrace{C_x(x,t)}_{MC} = 0$$

Note also that the cross partial derivative is given by

$$f_{xt}(x,t) = -C_{xt}(x,t).$$

Thus $f_{xt} \geq 0$ if and only if increasing t decreases marginal cost. It should not be surprising that this simple and intuitive condition is connected to monotone comparative statics on output: We would expect output to increase if marginal cost decreases for all levels of output. However, note that other restrictions are also required in order to apply the implicit function theorem method described above. For example, it relies on the assumption that the objective function is concave in x, that is,

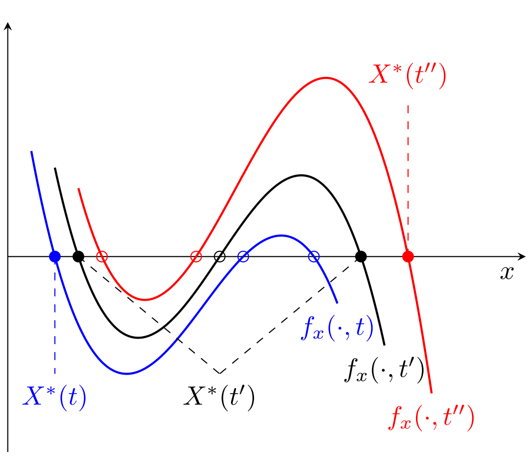
$$f_{xx}(x,t) = P''(x)x + 2P'(x) - C_{xx}(x,t) < 0.$$

When might this concavity condition be satisfied? Some sufficient conditions for concavity of the profit function are $P' \leq 0$, $P'' \leq 0$, and $C_{xx} > 0$, that is, the inverse demand function is decreasing and concave and the cost function is strictly convex.

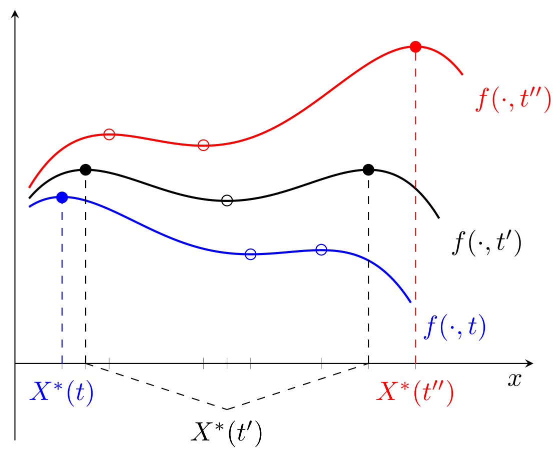
Unfortunately, there are some issues with the assumptions in the preceding example. Most notably, this example imposes strong and unrealistic assumptions about inverse demand. While it is plausible that inverse demand is decreasing, it is much less realistic to assume that it is concave. This also begs the question of whether the comparative statics in this example should depend on demand at all. Note that f_{xt} depends only on the cost function, and it seems quite plausible that monotone comparative statics could be obtained even without assuming that f is concave.

Let's explore relaxing some of the assumptions invoked in the implicit function theorem approach. Assume as before that $f_{xt} \geq 0$, but now relax the assumption that f is concave. Suppose that $x^*(t)$ is a critical point for f (that is, a solution to Equation (1.1)) but that

$f_{xx}(x^*(t), t) > 0$. In this case, Equation (1.2) implies that $x^{*\prime}(t) < 0$, which may give the impression that the comparative statics for maximizers of the function f are now reversed. However, that is not the case. Since $f_{xx}(x^*(t), t) > 0$, the second order condition for $x^*(t)$ to be a solution to the maximization problem is violated. Thus, while $x^*(t)$ is a critical point of the function f, it is *not* a solution to the maximization problem in this case. Figure 1.2 illustrates the partial derivative $f_x(\cdot, t)$ and the objective function $f(\cdot, t)$ for different values of t in the case where $f_{xt} > 0$ but f is not globally concave. Notice that the solutions (now there may be multiple solutions) are still increasing in an intuitive sense (that we will make precise shortly) even though the second-order condition is violated.



(a) Derivative f_x



(b) Objective function f

Figure 1.2: Solutions for the parameters $t < t' < t''$ in the non-concave case.

One might consider approaching the problem of comparative statics by extending the implicit function theorem method discussed above to deal with non-concave functions by formalizing the type of analysis illustrated in Figure 1.2. Instead, we will find that there is a much simpler and more powerful approach to comparative statics. In this chapter and the next, we will develop methods that relax not only concavity, but also differentiability and the assumption that the solution is interior.

1.2 Single Variable Comparative Statics

We begin with the one-dimensional case of $X \subseteq \mathbb{R}$ and $T \subseteq \mathbb{R}$. In Section 1.3, we will expand on this analysis to consider multidimensional choice variables and parameters.

1.2.1 Strong Set Order

Definition 1.2. For any $Y, Z \subseteq X$, we say that Z dominates Y in the **strong set order**, denoted $Z \geq_s Y$, if for every $y \in Y$ and $z \in Z$, $\min\{y, z\} \in Y$ and $\max\{y, z\} \in Z$.

- Equivalently, $Z \geq_s Y$ if for every $y \in Y$ and $z \in Z$,

$$y > z \implies y \in Z \text{ and } z \in Y.$$

- Does **not** require that for every $y \in Y$ and $z \in Z$, $z \geq y$.
- Illustrate (overlap permitted, but not $y > z$ and $y \notin Z$ or $z \notin Y$).
- E.g., for intervals $Y = [\underline{y}, \bar{y}]$ and $Z = [\underline{z}, \bar{z}]$, $Z \geq_s Y$ iff $\bar{z} \geq \bar{y}$ and $\underline{z} \geq \underline{y}$.
- Strong set order reduces to usual order for singletons: $\{z\} \geq_s \{y\}$ if and only if $z \geq y$.
- Empty set: $Y \geq_s \emptyset \geq_s Y$. (Useful: allows us to separate the study of comparative statics from the issue of existence of a solution.)

Lemma 1.3. Suppose $Y, Z \subseteq \mathbb{R}$, and suppose each of these sets has maximum and minimum elements, $\bar{y}, \underline{y} \in Y$ and $\bar{z}, \underline{z} \in Z$, respectively.[2] If $Z \geq_s Y$ then $\bar{z} \geq \bar{y}$ and $\underline{z} \geq \underline{y}$.

You are asked to prove this simple lemma in Exercise 1.1.

Definition 1.4. Suppose $X \subseteq \mathbb{R}$ and $T \subseteq \mathbb{R}$. A correspondence $X^* : T \twoheadrightarrow X$ is **monotone nondecreasing** in the strong set order if $t' \geq t$ implies $X^*(t') \geq_s X^*(t)$.

The following result follows immediately from Lemma 1.3. Note that compactness is only assumed in this result to ensure the existence of maximum and minimum elements of $X^*(t)$ for each t.

Corollary 1.5. Suppose $X \subseteq \mathbb{R}$, $T \subseteq \mathbb{R}$, and $X^* : T \twoheadrightarrow X$ is a nonempty-valued and compact-valued correspondence (i.e., $X^*(t)$ is nonempty and compact for all $t \in T$). Define functions $\bar{x}^* : T \to X$ and $\underline{x}^* : T \to X$ by $\bar{x}^*(t) = \max X^*(t)$ and $\underline{x}^*(t) = \min X^*(t)$. If X^* is monotone nondecreasing in the strong set order, then $\bar{x}^*$ and $\underline{x}^*$ are nondecreasing functions.

1.2.2 Increasing Differences

Definition 1.6. Suppose $X \subseteq \mathbb{R}$ and $T \subseteq \mathbb{R}$. A function $f : X \times T \to \mathbb{R}$ has **increasing differences** in $(x; t)$ if for all $x' > x$ and $t' > t$,

$$f(x', t') - f(x, t') \geq f(x', t) - f(x, t).$$

The condition above can also be written as

$$f(x', t') - f(x', t) \geq f(x, t') - f(x, t).$$

[2] That is, suppose there exist $\bar{y}, \underline{y} \in Y$ such that $\underline{y} \leq y \leq \bar{y}$ for all $y \in Y$, and likewise for Z. This assumption is satisfied, for example, if Y and Z are compact sets.

When $X \subseteq \mathbb{R}$ and $T \subseteq \mathbb{R}$, this property is equivalent to f being **supermodular** in (x,t), which is sometimes written as

$$f(x',t') + f(x,t) \geq f(x',t) + f(x,t').$$

However, the two concepts are not the same in general (when the parameter space or the decision variable space are multidimensional), as we will see later in Section 1.3.2.

The following lemma shows that in the case where f is differentiable with respect to x or t (or both), increasing differences boils down to some very simple conditions on the partial derivatives f_x and f_t, or on the cross-partial derivative f_{xt}.

Lemma 1.7. Suppose $X \subseteq \mathbb{R}$ and $T \subseteq \mathbb{R}$, and suppose $f : X \times T \to \mathbb{R}$. Consider the following set of statements:

(1) f has increasing differences in $(x;t)$.
(2) $f_x(x,t)$ is nondecreasing in t.
(3) $f_t(x,t)$ is nondecreasing in x.
(4) $f_{xt} \geq 0$.

These statements are related as follows:

- If X is an interval and $f_x(x,t)$ exists for all $x \in X$ and $t \in T$, then (1) $\iff$ (2).
- If T is an interval and $f_t(x,t)$ exists for all $x \in X$ and $t \in T$, then (1) $\iff$ (3).
- If both X and T are intervals and $f_{xt}(x,t)$ exists for all $x \in X$ and $t \in T$ (in particular, if f is twice continuously differentiable), then all four statements are equivalent.

In Exercise 1.2, you are asked to prove the equivalence of statements (1)–(4) in Lemma 1.7 in the case where X and T are intervals and under the assumption that f is twice continuously differentiable. This assumption is convenient as it ensures that all of the partial derivatives (and cross partial derivatives) are integrable and thus allows the result to be proved easily using the fundamental theorem of calculus. However, if you would like an added challenge, you can prove the lemma exactly as stated by instead using the mean value theorem.

We are now ready to state our first comparative statics result.

Theorem 1.8 (Topkis (1978)). Suppose $X \subseteq \mathbb{R}$, $T \subseteq \mathbb{R}$, and $f : X \times T \to \mathbb{R}$. If f has increasing differences in $(x;t)$, then $X^*(t) = \operatorname{argmax}_{x \in X} f(x,t)$ is monotone nondecreasing in t (in the strong set order), that is, $t' \geq t$ implies $X^*(t') \geq_s X^*(t)$.

Proof. Fix any $t' \geq t$, $x \in X^*(t)$, and $x' \in X^*(t')$. We need to show that if $x > x'$ then $x' \in X^*(t)$ and $x \in X^*(t')$. To see this holds under the stated assumptions, note that $x > x'$

implies

$$
\begin{aligned}
0 &\leq f(x,t) - f(x',t) && (x \in X^*(t)) \\
&\leq f(x,t') - f(x',t') && \text{(ID)} \\
&\leq 0. && (x' \in X^*(t'))
\end{aligned}
$$

Thus we must have $f(x,t) = f(x',t)$ and $f(x,t') = f(x',t')$, which implies $x' \in X^*(t)$ and $x \in X^*(t')$. $\square$

By Corollary 1.5, the largest and smallest solutions for each t (if they exist) are therefore nondecreasing functions. In particular, if $f(x,t)$ has a unique maximizer $x^*(t)$ for each t, then this solution is a nondecreasing function. Note that Theorem 1.8 not only relaxes the assumptions that f is differentiable and concave, it also does not require any assumptions about the solutions being in the interior of X. In fact, while the set X in this theorem could be an interval, it could alternatively be a discrete set of points or any other subset of the real line.

Example 1.9 (Monopoly response to change in cost). Consider again the problem of a firm choosing output to maximize profit:

$$f(x,t) = P(x)x - C(x,t)$$

If C has *decreasing* differences in $(x;t)$ (i.e., $-C$ has ID in $(x;t)$), then f has ID in $(x;t)$. For example, if C is twice continuously differentiable, this is equivalent to $C_{xt} \leq 0$. Thus, without any restrictions on the inverse demand function, we can conclude that output is nondecreasing (in the strong set order) in t, so long as increasing this parameter lowers marginal cost. In addition, Theorem 1.8 also allows us to easily deal with many nondifferentiable functions that arise naturally in economic applications. For example, suppose the cost for the firm can be decomposed into a variable cost and a fixed cost that is only incurred if the firm chooses to operate. That is, consider a cost function of the form

$$C(x,t) = \begin{cases} 0 & \text{if } x = 0 \\ F(t) + V(x,t) & \text{if } x > 0, \end{cases}$$

where $F(t)$ is the fixed cost and $V(x,t)$ is a variable cost function that is twice continuously differentiable and satisfies $V(0,t) = 0$. In this case, it is not difficult to show that $C(x,t)$ has decreasing differences in $(x;t)$ if and only if $F(t)$ is decreasing in t and $V_{xt} \leq 0$ (Exercise 1.9). Thus, if fixed cost and marginal cost are both decreasing in t, then an increase in t leads to higher production.

1.3 Multivariate Comparative Statics

1.3.1 Lattices and the Strong Set Order

For any vectors $x = (x_1, \ldots, x_n)$ and $x' = (x'_1, \ldots, x'_n)$ in $\mathbb{R}^n$, recall that we write $x' \geq x$ if $x'_i \geq x_i$ for all $i = 1, \ldots, n$. We write $x' > x$ if $x' \geq x$ and $x' \neq x$.

Definition 1.10. Let $x, x' \in \mathbb{R}^n$.

(1) The **meet** of x and x' is

$$x \wedge x' = (\min\{x_1, x'_1\}, \ldots, \min\{x_n, x'_n\}).$$

(2) The **join** of x and x' is

$$x \vee x' = (\max\{x_1, x'_1\}, \ldots, \max\{x_n, x'_n\})$$

- Illustrate meet and join for vectors in $\mathbb{R}$ and $\mathbb{R}^2$.
- Note that for any x, x', we have $x \wedge x' \leq x \leq x \vee x'$ and $x \wedge x' \leq x' \leq x \vee x'$.

Definition 1.11. A set $X \subseteq \mathbb{R}^n$ is a **lattice** if for all $x, x' \in X$, we have $x \wedge x' \in X$ and $x \vee x' \in X$. A set $Z \subseteq X$ is a **sublattice** of X if Z itself satisfies the definition of a lattice.

- Illustrate that a product set in $\mathbb{R}^2$ is a lattice.
- Illustrate that if a set in $\mathbb{R}^2$ is totally ordered (i.e., a chain) then it is automatically a lattice, as well as variations.
- A sublattice Z is simply a lattice that is a subset of another lattice X.

Definition 1.12. For any $Y, Z \subseteq X$, we say that Z dominates Y in the **strong set order**, denoted $Z \geq_s Y$, if for every $y \in Y$ and $z \in Z$, $y \wedge z \in Y$ and $y \vee z \in Z$.

- Note that reduces to previous definition for $n = 1$.
- Illustrate using a few examples.
- Note that $Z \geq_s Z$ if and only if Z is a sublattice of X.
- Mention projections onto each coordinate (Exercise 1.5).

Definition 1.13. A **partially ordered set** $(T, \geq)$ is a set T equipped with a partial order $\geq$. That is, the binary relation $\geq$ is:

(1) Transitive: $t \geq t'$ and $t' \geq t''$ imply $t \geq t''$.
(2) Reflexive: $t \geq t$.
(3) Antisymmetric: $t \geq t'$ and $t' \geq t$ imply $t = t'$.

We write $t > t'$ if $t \geq t'$ and $t' \not\geq t$. For this chapter, we will focus on $T \subseteq \mathbb{R}^m$ for concreteness. This is a partially ordered set when endowed with the usual order described above. However, the comparative statics results developed in this chapter also apply to more abstract settings (e.g., a set of probability distributions is a partially ordered set when endowed with one of the stochastic orders that we will introduce in later chapters). Finally, keep in mind that we will not require that T be a lattice. For example, it could be any subset of $\mathbb{R}^m$.

1.3.2 Increasing Differences and Supermodularity

Increasing differences is defined just as before, but keep in mind that we are now dealing with $X \subseteq \mathbb{R}^n$.

Definition 1.14. Suppose X is a lattice and T is a partially ordered set. A function $f : X \times T \to \mathbb{R}$ has **increasing differences** in $(x; t)$ if for all $x' > x$ and $t' > t$,

$$f(x', t') - f(x, t') \geq f(x', t) - f(x, t).$$

Lemma 1.15. If $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ is twice continuously differentiable, then f has increasing differences in $(x; t)$ if and only if $\partial^2 f / \partial x_i \partial t_j \geq 0$ for $i = 1, \ldots, n$, $j = 1, \ldots, m$.

EXAMPLE 1.16 (IS ID ENOUGH FOR MCS?). Change in choice of x and y for $2/5 < t < 4$.

- $f(x, y, t) = 3tx + (2 + t)y - (x + y)^2 - x^2 - y^2$.
- $f_x(x, y, t) = 3t - 4x - 2y, \quad f_{xt}(x, y, t) = 3$
- $f_y(x, y, t) = 2 + t - 2x - 4y, \quad f_{yt}(x, y, t) = 1$

$$\begin{aligned}
3t &= 4x + 2y \\
2 + t &= 2x + 4y \\
&\implies 6t - (2 + t) = 5t - 2 = 6x \implies x = (5t - 2)/6 \\
&\implies (4 + 2t) - 3t = 4 - t = 6y \implies y = (4 - t)/6
\end{aligned}$$

- Issue: $f_{xy}(x, y, t) = -2 < 0$
- Interpretation using two managers (one choosing x and the other choosing y) with the same objective of maximizing firm profits, but who only myopically best respond to the previous period output choice of the other manager: $\uparrow t \Rightarrow \uparrow (x, y) \Rightarrow \downarrow (x, y) \Rightarrow \uparrow (x, y) \Rightarrow \ldots$. If this sequence of responses converges, then the limit is the new solution for the new value of t. Note that we cannot make a definitive qualitative conclusion about the final change from such a sequence of adjustments since there are adjustments both up and down for x and y.

Definition 1.17. Suppose X is a lattice and T is a partially ordered set. A function $f : X \times T \to \mathbb{R}$ is **supermodular** in x if for all $x, x' \in X$ and $t \in T$,

$$f(x \wedge x', t) + f(x \vee x', t) \geq f(x, t) + f(x', t).$$

The condition above can equivalently be written as

$$f(x \vee x', t) - f(x', t) \geq f(x, t) - f(x \wedge x', t).$$

As was the case with increasing differences, there is also a useful characterization of supermodularity for differentiable functions.

Lemma 1.18. If $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ is twice continuously differentiable, then f is supermodular in x if and only if $\partial^2 f / \partial x_i \partial x_j \geq 0$ for $i, j = 1, \ldots, n$, $i \neq j$.

We say that f is **submodular** in x if the function $-f$ is supermodular in x. Note that an equivalent way to define of submodularity is to take the definition of supermodularity and reverse the inequality. By the preceding lemma, if f twice continuously differentiable, then it is submodular if and only if $\partial^2 f / \partial x_i \partial x_j \leq 0$ for all $i \neq j$.

We are now ready to establish our first multivariate comparative statics theorem.

Theorem 1.19 (Topkis (1978)). Suppose X is a lattice, T is a partially ordered set, and $f : X \times T \to \mathbb{R}$. If f is supermodular in x and has increasing differences in $(x; t)$, then $\operatorname{argmax}_{x \in X} f(x, t)$ is monotone nondecreasing in t (in the strong set order).

Proof. Let $X^*(t) = \operatorname{argmax}_{x \in X} f(x, t)$. Fix any $t' \geq t$, $x \in X^*(t)$, and $x' \in X^*(t')$. Note that

$$\begin{aligned} 0 &\leq f(x, t) - f(x \wedge x', t) && (x \in X^*(t)) \\ &\leq f(x \vee x', t) - f(x', t) && (\text{SM}) \\ &\leq f(x \vee x', t') - f(x', t') && (\text{ID}) \\ &\leq 0. && (x' \in X^*(t')) \end{aligned}$$

Thus $f(x, t) = f(x \wedge x', t)$ and $f(x \vee x', t') = f(x', t')$, which implies $x \wedge x' \in X^*(t)$ and $x \vee x' \in X^*(t')$. □

Table 1.1 illustrates Lemmas 1.15 and 1.18 in the case of a twice continuously differentiable function $f : X \times T \to \mathbb{R}$ for $X, T \subseteq \mathbb{R}^2$. Tables 1.1a and 1.1b show the implications of assuming supermodularity in x in addition to ID in $(x; t)$. Adding this assumption of SM in x rules out the type of objective function that appeared in Example 1.16 and (together with

	x_1	x_2	t_1	t_2
x_1			+	+
x_2			+	+
t_1	+	+		
t_2	+	+		

(a) ID in $(x;t)$

	x_1	x_2	t_1	t_2
x_1		+	+	+
x_2	+		+	+
t_1	+	+		
t_2	+	+		

(b) ID in $(x;t)$ and SM in x

	x_1	x_2	t_1	t_2
x_1		+	+	+
x_2	+		+	+
t_1	+	+		+
t_2	+	+	+	

(c) SM in (x,t) (not needed for MCS)

Table 1.1: Implications of increasing differences and supermodularity for cross-partial derivatives of twice continuously differentiable function $f : X \times T \to \mathbb{R}$ for $X, T \subseteq \mathbb{R}^2$.

ID) ensures monotone comparative statics of multidimensional choice variables with respect to increases in the parameter t, as Theorem 1.19 demonstrates.

It is important to note that the combined assumptions of ID in $(x;t)$ and SM in x are still weaker than assuming SM in *all* variables (x,t) simultaneously. Table 1.1c illustrates the implications of assuming SM in all variables, which is an *unnecessarily strong* restriction — we do not need $\partial^2 f/\partial t_1 \partial t_2 \geq 0$ for Theorem 1.19. This should be intuitive. We do not care about how the individual would like to adjust t_1 in response to t_2 (and vice versa) since neither of these variables are under the control of the individual. These parameters must be taken as given, and the individual is only free to choose x_1 and x_2.

We now consider a variation of Example 1.16 where the objective function f that has increasing differences in $(x;t)$ and is supermodular in x.

EXAMPLE 1.20. Consider the following variation of Example 1.16, where $\alpha \in (0,1)$ is a fixed constant (note that Example 1.16 considered the case of $\alpha = 2 > 1$):

- $f(x,y,t) = 3tx + (2+t)y - (x+y)^\alpha - x^2 - y^2$.
- $f_{xt}(x,y,t) = 3$.
- $f_{yt}(x,y,t) = 1$.
- $f_{xy}(x,y,t) = -\alpha(\alpha-1)(x+y)^{\alpha-2} \geq 0$ since $\alpha \in (0,1)$.
- The optimal x and y are nondecreasing in t when $\alpha \in (0,1)$.

1.4 Useful Techniques for Applications

In this section, we summarize some useful techniques developed in Milgrom and Shannon (1994) that can be used to apply monotone comparative statics to various economic problems.

1.4.1 Monotone Transformations

Suppose the function $f : X \times T \to \mathbb{R}$ fails to have increasing difference in $(x;t)$. However, suppose there exists a strictly increasing function $g : \mathbb{R} \to \mathbb{R}$ such that $g \circ f$ has increasing differences in $(x;t)$ (it suffices for g to have a domain that includes the range of f, not necessarily all of $\mathbb{R}$). Then since

$$\operatorname*{argmax}_{x \in X} f(x,t) = \operatorname*{argmax}_{x \in X} g(f(x,t)),$$

we can apply our comparative statics results to this transformed objective function.

Example 1.21 (Increasing elasticity of demand). Suppose a firm chooses price p to maximize profit

$$\pi(p,\alpha) = D(p,\alpha)(p-c),$$

where $c > 0$ is the constant marginal cost and $D(p,\alpha) = p^{-\alpha}$ is the demand function. Recall that this is the family of constant elasticity of demand functions:

$$\mathcal{E}(p,\alpha) = -\frac{p}{D(p,\alpha)} \frac{\partial D(p,\alpha)}{\partial p} = \alpha.$$

It can be shown that π does not have increasing differences in $(p;\alpha)$ or $(p;-\alpha)$ (Exercise 1.10). However, the monotone transformation

$$\begin{aligned}\log(\pi(p,\alpha)) &= \log(D(p,\alpha)) + \log(p-c) \\ &= -\alpha \log(p) + \log(p-c)\end{aligned}$$

does have increasing differences in $(p;-\alpha)$ (i.e., decreasing differences in $(p;\alpha)$). Note that this transformation is not well-defined for $p \leq c$. However, since any solution clearly must satisfy $p > c$, we can restrict attention to $p \in (c,\infty)$ without altering the solution set. Thus, the optimal price p is nonincreasing in α (in the strong set order). This conclusion should be intuitive. If demand is more elastic, then demand drops more quickly in response to an increase in price; hence the firm will find it optimal to set a lower price.

Notice that the conclusion of Example 1.21 can be generalized: If $D(p,\alpha)$ is a demand function and increasing α increases the elasticity of demand, then $\log(\pi(p,\alpha))$ has decreasing differences in $(p;\alpha)$.

1.4.2 Parameter-Contingent Transformations

In the previous section, we applied the same transformation g for all parameter values t. In fact, such independence is not required, and this transformation could also depend on the parameter. Formally, if $g : \mathbb{R} \times T \to \mathbb{R}$ is strictly increasing in its first argument for all t, then

$$\operatorname*{argmax}_{x \in X} f(x,t) = \operatorname*{argmax}_{x \in X} g(f(x,t),t)$$

for every $t \in T$. The follow example illustrates a parameter-contingent transformation.

EXAMPLE 1.22 (INCREASE IN MARKET SIZE). Consider the effects of an increase in the market size on monopoly price. Each consumer in this market has an inverse demand given by $P(q)$. If the number of customers N is exogenous, then firm's problem is to choose a quantity per customer q to maximize:

$$\pi(q, N) = NqP(q) - C(Nq).$$

Without some restrictions on demand, we cannot conclude that π has increasing differences in $(q; N)$. However, note that for $N > 0$, we have

$$\operatorname*{argmax}_{q\geq 0} \pi(q, N) = \operatorname*{argmax}_{q\geq 0} \frac{\pi(q, N)}{N} = \operatorname*{argmax}_{q\geq 0} \left(qP(q) - \frac{C(Nq)}{N} \right).$$

Thus, $q^*(N)$ is nondecreasing in N if $-C(Nq)/N$ has increasing differences in $(q; N)$. Assuming that C is twice continuously differentiable (this assumption is not required for the conclusion but makes the characterization of increasing differences simpler),

$$\frac{d^2(-C(Nq)/N)}{dNdq} = \frac{d(-C'(Nq))}{dN} = -qC''(Nq).$$

Thus $q^*(N)$ is nondecreasing in N if C is concave, and $q^*(N)$ is nonincreasing in N if C is convex. No restrictions on the inverse demand function are required for this conclusion.

It is important to note that taking a transformation that depends on the choice variable x would alter the set of maximizers. For example, dividing by q in the previous example would change the solution set $q^*(N)$ and could therefore yield incorrect conclusions. Be careful when applying this method that your transformation only involves parameters and not choice variables.

1.4.3 Aggregation

EXAMPLE 1.23 (RESPONSE TO A PRICE CHANGE). Consider a firm that produces a quantity x of an output using a vector $z \in \mathbb{R}^k$ of inputs. If the input and output markets are competitive, then the firm acts as a price-taker. Given a production function $F : \mathbb{R}^k_+ \to \mathbb{R}_+$, output price $p > 0$, and input prices $w \in \mathbb{R}^k_{++}$, the firm's problem is thus

$$\max_{\substack{x\in\mathbb{R}_+, z\in\mathbb{R}^k_+ \\ x\leq F(z)}} \big(px - w \cdot z\big) = \max_{z\in\mathbb{R}^k_+} \big(pF(z) - w \cdot z\big).$$

Suppose we are interested in how output responds to an increase in the price p. To determine this relationship, we can first separate the problem into one of cost minimization followed by selection of profit-maximizing output:

$$\pi(x, p) = px - C(x),$$

where

$$C(x) = \min\{w \cdot z : z \in \mathbb{R}^k, x \leq F(z)\}.$$

The profit function π has increasing differences in $(x; p)$, and hence profit-maximizing output is nondecreasing in p. No restrictions on the production technology are required for this conclusion.

1.4.4 Composite Functions

Example 1.24 (Choosing market size). Suppose the monopoly in Example 1.22 is able to choose its market size N, for example, by adding new outlets or by increasing advertising. If the firm finds a new and less expensive way of expanding its market, how would this influence its choice of market size N and quantity per customer q? Formally, suppose the firm chooses (q, N) to maximize $NqP(q) - C(Nq) - K(N, t)$ where K is submodular ($K_{Nt} \leq 0$). Using the aggregation method, we can write the problem of selecting a market size N as

$$\max_{N>0} \Big(g(N) - K(N, t)\Big),$$

where

$$g(N) = \max_{q \geq 0} \Big(NqP(q) - C(Nq)\Big).$$

Since $g(N) - K(N, t)$ has increasing differences in $(N; t)$ by the assumption that $K_{Nt} \leq 0$, we know that $N^*(t)$ is nondecreasing in t. Note that the optimal q depends only on N, and we know from Example 1.22 that $q^*(N)$ is nondecreasing in N if C is concave (and nonincreasing if C is convex). Thus $q^*(N^*(t))$ is nondecreasing in t if C is concave (and nonincreasing if C is convex). There are a few details left to check about how monotonicity in the strong set order is preserved under this composition of two correspondences. You are asked to complete the missing arguments in Exercise 1.11.

1.5 Exercises

1.1 In this problem, you are asked to prove that $Z \geq_s Y$ implies that the greatest and least elements of Z and Y are ordered.

(a) Prove Lemma 1.3.

(b) Is the converse of Lemma 1.3 also true? That is, does $\bar{z} \geq \bar{y}$ and $\underline{z} \geq \underline{y}$ imply $Z \geq_s Y$? Prove or provide a counterexample.

1.2 Prove the equivalence of statements (1)–(4) in Lemma 1.7 in the case where X and T are intervals and f is twice continuously differentiable. (Hint: Use the fundamental theorem of calculus.) You can also prove the actual statement of the lemma for an added challenge (hint: mean value theorem), but that is not required.

1.3 Consider a one-period work-consumption problem. The individual has two sources of consumption: accumulated wealth and income earned in the current period. Her utility function for $(i, w) \in \mathbb{R}_+^2$ takes the form:

$$U(i, w) = u(w + i) - e(i),$$

where $u : \mathbb{R}_+ \to \mathbb{R}$ is her utility for money and $e : \mathbb{R}_+ \to \mathbb{R}$ is the effort cost associated with earning income i.[3] The individual takes w as given can chooses i optimally. Suppose u is twice continuously differentiable and $u'' \leq 0$.

(a) Does U have increasing differences $(i; w)$? If not, is there a way to transform this problem into one that does have increasing differences?

(b) What conclusions can be made about the relationship between w and the optimal income i?

1.4 Consider a two-period consumption-savings problem. The suppose the individual has the following utility function for $(c_1, c_2) \in \mathbb{R}_+^2$:

$$U(c_1, c_2) = u(c_1) + \beta u(c_2),$$

where $\beta \in (0, 1)$ and $u : \mathbb{R}_+ \to \mathbb{R}$ is strictly increasing. The individual has initial wealth w at the start of period 1 and can save between the two periods at a (deterministic) gross interest rate of $R > 0$.

(a) Suppose you know that u is twice continuously differentiable and concave, $u'' \leq 0$. What conclusions can be made about how the optimal c_1 and c_2 change with wealth w?[4]

(b) Suppose we relax the differentiability assumption from part (a). That is, do not make any assumptions about whether u is differentiable, but maintain the assumption that u is concave. Under these weaker assumptions, what conclusions can be made about how the optimal c_1 and c_2 change with w?[5]

(c) In the previous parts of the problem, is it possible to have a selection $c_1^*(w)$ from the optimal consumption choices in period 1 that strictly decreases in w at some wealth levels, so $w' > w$ and $c_1^*(w') < c_1^*(w)$? Either provide an example where this can happen or prove that it cannot happen. Also, if your answer is that $c_1^*(w)$ can strictly decrease in w under the assumptions given in this problem, what change in the assumptions would ensure that every selection $c_1^*(w)$ from the

[3] Think of this as an indirect "utility cost" function that minimizes the effort cost needed to earn a particular income i, taking into account the wages and utility costs associated with one, or perhaps many, uses of labor effort. This is analogous to the cost minimization problem of a firm.

[4] Hint: It may be helpful to substitute out for c_2 and express utility as a function of c_1 and w. Then do a similar substitution for c_1.

[5] Hint: If u is concave, then for any $a' > a$ and $b > 0$, $u(a + b) - u(a) \geq u(a' + b) - u(a')$. You can use this fact without proving it, but it is another good exercise to prove that this is true for any concave function, whether or not it is differentiable.

solution is monotonically nondecreasing in w?

1.5 Suppose $X \subseteq \mathbb{R}^n$ is a lattice. We have shown conditions under which the solution to a parameterized maximization problem $X^*(t)$ is nondecreasing in the strong set order (on $\mathbb{R}^n$). However, we are often interested in whether the set of solutions for one of the variables $X_i^*(t)$ for $i \in \{1, \ldots, n\}$ is nondecreasing in the strong set order on $\mathbb{R}$. Fix any $Y, Z \subseteq X$. For each i define

$$Y_i = \{a \in \mathbb{R} : \exists y \in Y \text{ such that } y_i = a\}.$$

and define Z_i similarly. In other words, Y_i and Z_i are the projections of the sets Y and Z, respectively, onto the ith coordinate.

(a) Suppose $Z \geq_s Y$ in the strong set order on $\mathbb{R}^n$. Is $Z_i \geq_s Y_i$ in the strong set order on $\mathbb{R}$? Prove or provide a counterexample.

(b) Suppose $Z_i \geq_s Y_i$ in the strong set order on $\mathbb{R}$ for each $i \in \{1, \ldots, n\}$. Is $Z \geq_s Y$ in the strong set order on $\mathbb{R}^n$? Prove or provide a counterexample.

1.6 Suppose that $X, Y, Z \subseteq \mathbb{R}^n$ are nonempty and $Z \geq_s Y$ and $Y \geq_s X$. Prove that $Z \geq_s X$. That is, the strong set order is transitive for nonempty sets.

1.7 Answer the following:

(a) Suppose $X \subseteq \mathbb{R}^n$ is a lattice, and $f : X \to \mathbb{R}$ and $g : X \to \mathbb{R}$ are supermodular. For any $\alpha \in (0, 1)$, the convex combination of f and g, denoted $\alpha f + (1 - \alpha)g$, is defined by

$$(\alpha f + (1 - \alpha)g)(x) = \alpha f(x) + (1 - \alpha)g(x).$$

Is the convex combination of supermodular functions supermodular? Prove or provide a counterexample.

(b) Suppose X_1, X_2 and X_3 are subsets of $\mathbb{R}$ and $f : X_1 \times X_2 \times X_3 \to \mathbb{R}$ is supermodular function. Define $F(x_1, x_2) = \max_{x_3 \in X_3} f(x_1, x_2, x_3)$ (assume the maximizer exists). Is F supermodular? Prove or provide a counterexample.

1.8 Consider the effects of an increase in the market size on the decisions of a monopoly firm. Each consumer in the market has an inverse demand given by $P(q)$. If the number of customers $N \geq 0$ is exogenous, then the firm's problem is to choose a quantity per customer $q \geq 0$ to maximize:

$$\pi(q, N) = NqP(q) - C(Nq),$$

where C is the firm's cost function. Assume the inverse demand function takes the parametric form

$$P(q) = q^{-\alpha},$$

for some $\alpha \in (0,1)$, and the cost function takes the parametric form

$$C(q) = q + \beta \exp(\beta\delta q) - \beta,$$

for some $\beta \in \mathbb{R}$ and $\delta > 0$. Answer the following. Remember that you must provide arguments to support all of your answers, and you will not receive full credit if you include any unnecessary parameter restrictions.

(a) Starting with some basics:
  i. Is $C(q)$ nonnegative and nondecreasing in $q \geq 0$ for all of the permitted parameter values (permitted meaning that β can be any real number, but δ must be strictly positive)?
  ii. For which parameter values is C convex, and for which is it concave?

(b) Under what, if any, restrictions on the parameters α, β, and/or δ is the optimal quantity per customer q nondecreasing (in the strong set order) in N? Under what restrictions is q nonincreasing in N?

(c) Under what, if any, restrictions on the parameters α, β, and/or δ is the optimal total quantity $Q \equiv Nq$ produced by the firm nondecreasing (in the strong set order) in N?

(d) Under what, if any, restrictions on the parameters α, β, and/or δ is the optimal price charged by the monopolist nondecreasing (in the strong set order) in N?

(e) Under what, if any, restrictions on the parameters α, β, and/or δ is the profit of the monopolist nondecreasing in N?

1.9 Suppose as in Example 1.9 that $X = \mathbb{R}_+$, $T \subseteq \mathbb{R}$ is an interval, and $C : X \times T \to \mathbb{R}$ is defined by

$$C(x,t) = \begin{cases} 0 & \text{if } x = 0 \\ F(t) + V(x,t) & \text{if } x > 0, \end{cases}$$

where $F(t)$ is the fixed cost and $V(x,t)$ is a variable cost function that is twice continuously differentiable and satisfies $V(0,t) = 0$. Show that $C(x,t)$ has decreasing differences in $(x;t)$ if and only if $F(t)$ in nonincreasing in t and $V_{xt} \leq 0$.

1.10 Suppose as in Example 1.21 that $\pi(p,\alpha) = D(p,\alpha)(p-c)$ where $D(p,\alpha) = p^{-\alpha}$. Show that π has neither increasing nor decreasing differences in $(p;\alpha)$.

1.11 Suppose $T \subseteq \mathbb{R}$, $X \subseteq \mathbb{R}$, and consider two correspondences $\varphi : T \twoheadrightarrow X$ and $\psi : X \twoheadrightarrow \mathbb{R}$. Define the composition of these correspondences as follows:

$$\psi(\varphi(t)) \equiv \{y \in \mathbb{R} : y \in \psi(x) \text{ for some } x \in \varphi(t)\} = \bigcup_{x \in \varphi(t)} \psi(x).$$

If $\varphi(t)$ and $\psi(x)$ are both monotone nondecreasing in the strong set order, then is the composition $\psi(\varphi(t))$ monotone nondecreasing in the strong set order? Prove or provide a counterexample.

Chapter 2

Ordinal Theory of Monotone Comparative Statics

Contents

References and Assigned Readings

Primary readings:

(1) Milgrom and Shannon (1994) (recommended).
(2) Dekel notes on Monotone Comparative Statics (optional).
(3) Milgrom and Roberts (1996) (required).

Additional references:

(1) Topkis (1998)

2.1 Single Variable Comparative Statics

As in the previous chapter, we begin our analysis with the case of $X \subseteq \mathbb{R}$ and $T \subseteq \mathbb{R}$.

2.1.1 Single Crossing Property

In Sections 1.4.1 and 1.4.2, we introduced the technique of taking monotone transformations of the objective function. The invariance of the solution set to such monotone transformations illustrates that monotonicity of the solution set follows from the ordinal properties of the objective function. That is, knowing the level curves of $f(\cdot, t)$ for each t is sufficient to determine the set of solutions and how they vary with t. For a familiar analogy, since utility functions are only identified up to a monotone transformations, we often focus on indifference curves and ordinal properties (e.g., quasiconcavity) rather than cardinal properties (e.g., concavity). Similarly, the single crossing property is an ordinal condition that relaxes the cardinal property of increasing differences.

Definition 2.1. Suppose $X \subseteq \mathbb{R}$ and $T \subseteq \mathbb{R}$. A function $f : X \times T \to \mathbb{R}$ satisfies the **single crossing property** in $(x; t)$ if for all $x' > x$ and $t' > t$,

$$f(x', t) \geq f(x, t) \implies f(x', t') \geq f(x, t')$$

and

$$f(x', t) > f(x, t) \implies f(x', t') > f(x, t').$$

The single crossing property is named after the similar property of a single-variable function. Fix any $x' > x$ and define a function $g : T \to \mathbb{R}$ by $g(t) = f(x', t) - f(x, t)$. The single crossing property implies that g crosses zero at most once and from below.[1] Equivalently,

[1]However, there could be an interval of parameter values where $g(t) = 0$. For example, it could be that $g(t) = g(t') = 0$ for some $t' > t$. In this case, the single crossing property requires that $g(t'') = 0$ for all $t'' \in (t, t')$. Similarly, a function $f : X \times T \to \mathbb{R}$ that has the single crossing property in $(x; t)$ could have $f(x', t) = f(x, t)$ and $f(x', t') = f(x, t')$ for $x' > x$ and $t' > t$, so long as $f(x', t'') = f(x, t'')$ for all $t'' \in (t, t')$.

for fixed $x' > x$, the function $f(x', \cdot)$ crosses the function $f(x, \cdot)$ at most once and from below. Keep in mind that to check for the single crossing property, we want to look for single crossing of the function as we increase t while holding fixed the pair of values x and x'.[2]

Several important features of the single crossing property (SC) and its connection with increasing differences (ID) are worth highlighting:

(1) If f has increasing differences in $(x; t)$, then it has the single crossing property in $(x; t)$. That is, ID implies SC.

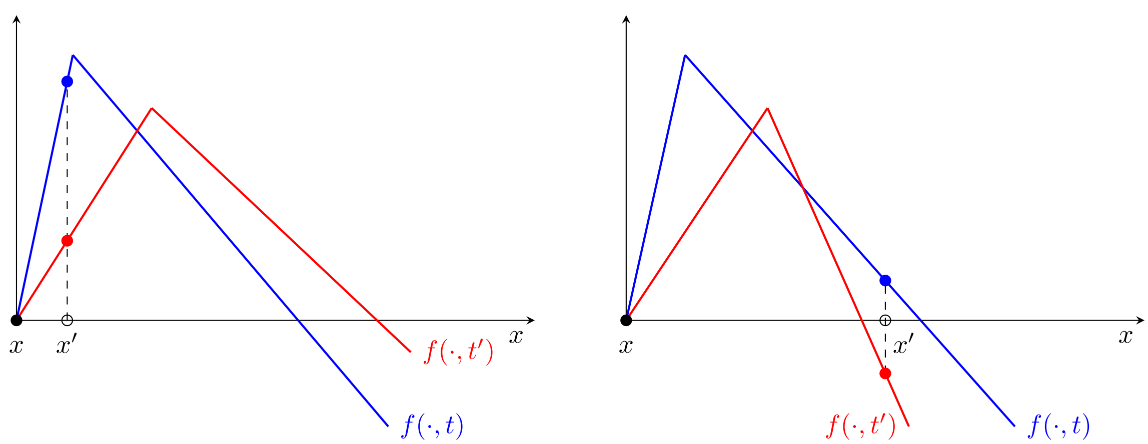
(2) If f has SC in $(x; t)$, then so does any strictly increasing transformation of f (as in Section 1.4.1) or any parameter-contingent transformation of f (as in Section 1.4.2).

Both observations follow directly from the definitions of SC and ID. The first observation implies that SC is a weaker (less restrictive) condition that ID. The second observation implies that SC is an ordinal property: It is robust to monotone transformations of the objective function. Combining these two observations, we obtain the following useful sufficient condition that can sometimes be used to verify that a function has the single cross property:

(3) If a strictly increasing transformation of f has ID, then f has SC.

Note that this condition is sufficient for SC, but not necessary. You are asked to explore the relationship between ID and SC and the role of monotone transformations in more detail in Exercises 2.4 and 2.5.

Figure 2.1 illustrates the distinction between increasing differences and the single crossing property: Figure 2.1a shows a function that violates increasing differences, but satisfies the single crossing property. Figure 2.1b illustrates a function that violates both ID and SC.



(a) f violates ID, but satisfies SC

(b) f violates both ID and SC

Figure 2.1: The single crossing property is weaker than increasing differences ($t < t'$).

[2]In particular, the single crossing property in $(x; t)$ is *not* equivalent to saying that for $t' > t$, the function $f(\cdot, t')$ crosses $f(\cdot, t)$ once from below—this would be the single crossing property in $(t; x)$.

The following is our first main result for single variable comparative statics.

Theorem 2.2 (Milgrom and Shannon (1994)). Suppose $X \subseteq \mathbb{R}$, $T \subseteq \mathbb{R}$, and $f : X \times T \to \mathbb{R}$. If f satisfies the single crossing property in $(x;t)$, then $\operatorname{argmax}_{x \in X} f(x,t)$ is monotone nondecreasing in t (in the strong set order).

Proof. Let $X^*(t) = \operatorname{argmax}_{x \in X} f(x,t)$. Fix any $t' \geq t$ and $x \in X^*(t)$, $x' \in X^*(t')$. We need to show that if $x > x'$ then $x' \in X^*(t)$ and $x \in X^*(t')$. To see this, suppose $x > x'$, and note first that

$$\begin{aligned} x \in X^*(t) &\implies f(x,t) \geq f(x',t) \\ &\implies f(x,t') \geq f(x',t') \qquad \text{(SC)} \\ &\implies x \in X^*(t'). \end{aligned}$$

Next, we prove that $x' \in X^*(t)$ by contradiction:

$$\begin{aligned} x' \notin X^*(t) &\implies f(x,t) > f(x',t) \\ &\implies f(x,t') > f(x',t') \qquad \text{(SC)} \\ &\implies x' \notin X^*(t'), \end{aligned}$$

a contradiction to the assumption that $x' \in X^*(t')$. Thus we must have $x' \in X^*(t)$. $\square$

2.1.2 When Single Crossing Cannot Be Weakened

One may wonder whether there is a condition even weaker than the single crossing property that can be used to obtain monotone comparative statics. It turns out that the answer depends on the flexibility that we have in specifying the constraint set. If the constraint set for x is fixed at X, then a condition like single crossing may not be necessary for solutions to be nondecreasing in the parameter t (although SC is of course sufficient for this). On the other extreme, if we require monotonicity of the solution set in t for all possible constraint sets $S \subseteq X$, then the single crossing property becomes a necessary condition. In fact, as the following theorem shows, we only need monotonicity for each binary set $S \subseteq X$, that is, each set of the form $S = \{\bar{x}, \bar{x}'\}$.

Theorem 2.3. Suppose $X \subseteq \mathbb{R}$, $T \subseteq \mathbb{R}$, and $f : X \times T \to \mathbb{R}$. If $\operatorname{argmax}_{x \in S} f(x,t)$ is monotone nondecreasing in t (in the strong set order) for each $S \subseteq X$ of the form $S = \{\bar{x}, \bar{x}'\}$, then f satisfies the single crossing property in $(x;t)$.

Proof. Fix any $\bar{x}' > \bar{x}$ and let $S = \{\bar{x}, \bar{x}'\}$. Let $X^*(t) = \operatorname{argmax}_{x \in S} f(x,t)$. Then, for any

$t' > t$, we have

$$\begin{aligned} f(\bar{x}', t) \geq f(\bar{x}, t) &\implies \bar{x}' \in X^*(t) \\ &\implies \bar{x}' \in X^*(t') && (X^* \text{ is nondecreasing}) \\ &\implies f(\bar{x}', t') \geq f(\bar{x}, t'). \end{aligned}$$

Similarly,

$$\begin{aligned} f(\bar{x}', t) > f(\bar{x}, t) &\implies \bar{x} \notin X^*(t) \\ &\implies \bar{x} \notin X^*(t') && (X^* \text{ is nondecreasing}) \\ &\implies f(\bar{x}', t') > f(\bar{x}, t'). \end{aligned}$$

Thus, f satisfies the single crossing property. □

How do we connect these necessary and sufficient conditions? Recall that Theorem 2.2 states that if f has SC in $(x;t)$, then the set of maximizers in $X \subseteq \mathbb{R}$ is nondecreasing in t. Of course, by a simple relabeling, this means that the set of all maximizers from any set $S \subseteq X$ is also nondecreasing in t when f has SC. In particular, this implies that the same is true for any binary set S. Finally, Theorem 2.3 states that monotonicity of the solution in t for every binary constraint set S implies that f has SC. The diagram in Figure 2.2 illustrates these logical relationships.

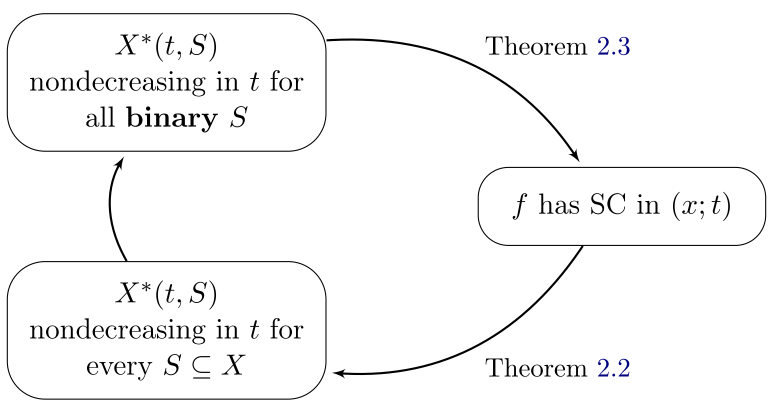


Figure 2.2: Relationship between properties of the function f and properties of the solution set $X^*(t, S) = \text{argmax}_{x \in S} f(x, t)$ in the single-dimensional case ($X \subseteq \mathbb{R}$).

Combining these observations, we obtain the following corollary.

Corollary 2.4. Suppose $X \subseteq \mathbb{R}$, $T \subseteq \mathbb{R}$, and $f : X \times T \to \mathbb{R}$. Then $\text{argmax}_{x \in S} f(x, t)$ is monotone nondecreasing in t (in the strong set order) for each $S \subseteq X$ if and only if f has the single crossing property in $(x;t)$.

Corollary 2.4 implies that the single crossing property is the weakest condition that can ensure monotone comparative statics on every possible constraint set $S \subseteq X$.[3]

2.1.3 When Increasing Differences Cannot Be Weakened

We have now observed that increasing differences is sufficient to obtain monotone comparative statics, but it is not necessary. In particular, the single crossing property is a weaker condition that is also sufficient for comparative statics. However, interestingly enough, there are circumstances where ID might become a necessary condition. Suppose the objective function takes the form

$$g(x,t) = f(x,t) - px.$$

For example, if f is the revenue function for a firm (output level times a fixed output price) as a function of the quantity x of an input and if p is the unit price of that input, then g is the profit function for the firm. It is easy to see that if the function f has ID, then so does g. But if the function f only satisfies SC, then g may fail to satisfy SC. Figure 2.3 illustrates this possibility using the example introduced previously in Figure 2.1a. The following theorem shows that if all values of p are possible in this objective function, then the only way to ensure that g has SC is if f has ID.

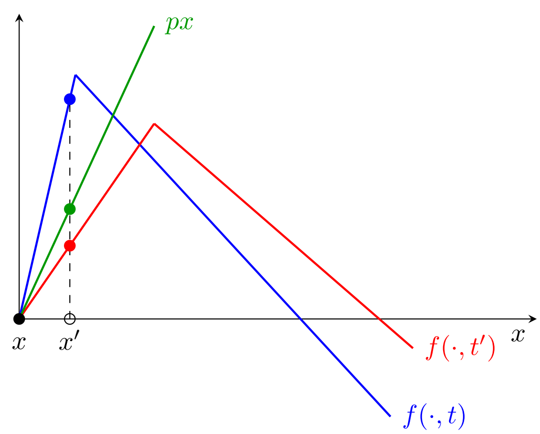


Figure 2.3: The function $f(x,t)$ satisfies SC, but $f(x,t) - px$ violates SC.

Theorem 2.5. Let $X \subseteq \mathbb{R}$, $T \subseteq \mathbb{R}$, and $f : X \times T \to \mathbb{R}$. Then $f(x,t) - px$ has the single crossing property in $(x;t)$ for all $p \in \mathbb{R}$ if and only if f has increasing differences in $(x;t)$.

Proof. If f has increasing differences in $(x;t)$, then $f(x,t) - px$ also has increasing differences in $(x;t)$ and therefore $f(x,t) - px$ has the single crossing property in $(x;t)$.

To prove the converse, suppose f does not have increasing differences in $(x;t)$. Then there

[3]What if we only need MCS for some smaller class of subsets such as intervals? In this case, we can use a weaker condition, such as the interval dominance order (Quah and Strulovici (2009)).

exist $x' > x$ and $t' > t$ such that

$$f(x', t') - f(x, t') < f(x', t) - f(x, t).$$

Choose $p \in \mathbb{R}$ such that

$$\frac{f(x', t') - f(x, t')}{x' - x} < p < \frac{f(x', t) - f(x, t)}{x' - x}.$$

Rearranging these inequalities yields

$$f(x', t') - px' < f(x, t') - px \qquad \text{and} \qquad f(x', t) - px' > f(x, t) - px,$$

violating the single crossing property of $f(x, t) - px$. □

In the case where f is nondecreasing in x, Theorem 2.5 continues to hold if we replace $p \in \mathbb{R}$ with the weaker restriction $p \geq 0$. It should be obvious from the proof why monotonicity of f in x permits this relaxation of the assumptions.

Combining Theorem 2.5 and Corollary 2.4, we obtain the following result.

Corollary 2.6. Suppose $X \subseteq \mathbb{R}$, $T \subseteq \mathbb{R}$, and $f : X \times T \to \mathbb{R}$. Then $\operatorname{argmax}_{x \in S}[f(x, t) - px]$ is monotone nondecreasing in t (in the strong set order) for each $S \subseteq X$ and for all $p \in \mathbb{R}$ if and only if f has increasing differences in $(x; t)$.

2.1.4 Strict Single Crossing

In this section, we explore a strong version of the single crossing property that ensures that any selection function from the solution correspondence is nondecreasing. You are asked to complete the proofs of these results in Exercise 2.2.

Definition 2.7. Suppose $X \subseteq \mathbb{R}$ and $T \subseteq \mathbb{R}$. A function $f : X \times T \to \mathbb{R}$ satisfies the **strict single crossing property** in $(x; t)$ if for all $x' > x$ and $t' > t$,

$$f(x', t) \geq f(x, t) \implies f(x', t') > f(x, t').$$

Theorem 2.8. Suppose $X \subseteq \mathbb{R}$, $T \subseteq \mathbb{R}$, and $f : X \times T \to \mathbb{R}$. Let $X^*(t) = \operatorname{argmax}_{x \in X} f(x, t)$. If f satisfies the strict single crossing property in $(x; t)$, then for any $t' > t$, $x \in X^*(t)$, and $x' \in X^*(t')$, we have $x' \geq x$.

Corollary 2.9. Suppose $X \subseteq \mathbb{R}$, $T \subseteq \mathbb{R}$, and $f : X \times T \to \mathbb{R}$. If f satisfies the strict single crossing property in $(x; t)$, then every selection $x^*(t)$ from $\operatorname{argmax}_{x \in X} f(x, t)$ must be nondecreasing in t.

Note that this result does not imply that every (or even any) selection from the solution correspondence is strictly increasing.

2.2 Multivariate Comparative Statics

2.2.1 Single Crossing Property and Quasisupermodularity

The definition of the single crossing property extends to any lattice $X \subseteq \mathbb{R}^n$ and any partially ordered set T (such as $T \subseteq \mathbb{R}^m$) without change.

Definition 2.10. Suppose X is a lattice and T is a partially ordered set. A function $f : X \times T \to \mathbb{R}$ satisfies the **single crossing property** in $(x;t)$ if for all $x' > x$ and $t' > t$,

$$f(x',t) \geq f(x,t) \implies f(x',t') \geq f(x,t')$$

and

$$f(x',t) > f(x,t) \implies f(x',t') > f(x,t').$$

Just as increasing differences can be relaxed to single crossing, supermodularity can be relaxed to quasisupermodularity.

Definition 2.11. Suppose X is a lattice and T is a partially ordered set. A function $f : X \times T \to \mathbb{R}$ is **quasisupermodular** in x if for all x, x' and for all t,

$$f(x,t) \geq f(x \wedge x',t) \implies f(x \vee x',t) \geq f(x',t)$$

and

$$f(x,t) > f(x \wedge x',t) \implies f(x \vee x',t) > f(x',t).$$

Theorem 2.12. Let X be a lattice, T be a partially ordered set, and $f : X \times T \to \mathbb{R}$. If f is quasisupermodular in x and satisfies the single crossing property in $(x;t)$, then $\operatorname{argmax}_{x\in X} f(x,t)$ is monotone nondecreasing in t (in the strong set order).

Proof. Let $X^*(t) = \operatorname{argmax}_{x\in X} f(x,t)$. Fix any $t' \geq t$, $x \in X^*(t)$, and $x' \in X^*(t')$. Note that

$$\begin{aligned} x \in X^*(t) &\implies f(x,t) \geq f(x \wedge x',t) \\ &\implies f(x \vee x',t) \geq f(x',t) && \text{(QSM)} \\ &\implies f(x \vee x',t') \geq f(x',t') && \text{(SC)} \\ &\implies x \vee x' \in X^*(t'). \end{aligned}$$

Next, we prove that $x \wedge x' \in X^*(t)$ by contradiction:

$$\begin{aligned} x \wedge x' \notin X^*(t) &\implies f(x,t) > f(x \wedge x', t) \\ &\implies f(x \vee x', t) > f(x', t) \qquad \text{(QSM)} \\ &\implies f(x \vee x', t') > f(x', t') \qquad \text{(SC)} \\ &\implies x' \notin X^*(t'), \end{aligned}$$

a contradiction to the assumption that $x' \in X^*(t')$. Therefore, it must be the case that $x \wedge x' \in X^*(t)$. □

2.2.2 Changing the Constraint Set

Suppose we consider the possibility of altering the constraint set instead of the parameter t, or we consider altering both the constraint set and the parameter. In the following theorem, we consider replacing the constraint set $S \subseteq X$ with another set $S' \subseteq X$ that dominates S in the strong set order.

Theorem 2.13. Let X be a lattice, T be a partially ordered set, and $f : X \times T \to \mathbb{R}$. If f is quasisupermodular in x and satisfies the single crossing property in $(x; t)$, then $\operatorname{argmax}_{x \in S} f(x,t)$ is monotone nondecreasing in t and S (in the strong set order), that is, for $t' \geq t$ and $S' \geq_s S$,

$$\operatorname*{argmax}_{x \in S'} f(x,t') \geq_s \operatorname*{argmax}_{x \in S} f(x,t).$$

Proof. Let $X^*(t,S) = \operatorname{argmax}_{x \in S} f(x,t)$. Fix any $t' \geq t$ and $S' \geq_s S$, and take any $x \in X^*(t,S)$ and $x' \in X^*(t',S')$. Since $x \in S$, $x' \in S'$, and $S' \geq_s S$, the definition of the strong set order states that $x \wedge x' \in S$ and $x \vee x' \in S'$. Next, the argument that $f(x \vee x', t') = f(x', t')$ and $f(x,t) = f(x \wedge x', t)$ is exactly the same as in the proof of Theorem 2.12. Combining these observations, conclude that $x \wedge x' \in X^*(t,S)$ and $x \vee x' \in X^*(t',S')$. □

In particular, suppose we have a constraint set $S(t)$ that is also parameterized by t, and suppose this set is monotone nondecreasing in t in the strong set order. Then Theorem 2.13 can be applied to show that under single crossing and quasisupermodularity, the solution set is nondecreasing in t, that is, for all $t' \geq t$,

$$\operatorname*{argmax}_{x \in S(t')} f(x,t') \geq_s \operatorname*{argmax}_{x \in S(t)} f(x,t).$$

2.2.3 Necessity of the Conditions

Theorem 2.3 showed the necessity of single crossing for monotonicity of the solutions in the case of a one-dimensional choice variable and parameter. The result extends immediately to any lattice X and partially ordered set T. The proof of the following result is literally identical to the proof of Theorem 2.3 for these more general spaces.

Theorem 2.14. Let X be a lattice, T be a partially ordered set, and $f : X \times T \to \mathbb{R}$. If $\text{argmax}_{x \in S} f(x,t)$ is monotone nondecreasing in t (in the strong set order) for each $S \subseteq X$ of the form $S = \{\bar{x}, \bar{x}'\}$ for $\bar{x}' > \bar{x}$, then f satisfies the single crossing property in $(x;t)$.

The following result shows that monotonicity of the solution with respect to changes in the constraint set implies quasisupermodularity. Since quasisupermodularity only plays a role for multi-dimensional choice variables (it is trivially satisfied for any single-dimensional choice variable), there is no analogue to this result from our study of single variable comparative statics.

Theorem 2.15. Let X be a lattice, T be a partially ordered set, and $f : X \times T \to \mathbb{R}$. Suppose $\text{argmax}_{x \in S} f(x,t)$ is monotone nondecreasing in S (in the strong set order) for each t, that is, $S' \geq_s S$ implies

$$\operatorname*{argmax}_{x \in S'} f(x,t) \geq_s \operatorname*{argmax}_{x \in S} f(x,t).$$

Then f is quasisupermodular in x.

Proof. Let $X^*(t,S) = \text{argmax}_{x \in S} f(x,t)$. Fix any t and $x, x' \in X$. Let $S = \{x \wedge x', x\}$ and $S' = \{x', x \vee x'\}$. Thus $S' \geq_s S$. Since X^* is nondecreasing in S, it follows that $x \in X^*(t,S)$ implies $x \vee x' \in X^*(t,S')$, and $x' \in X^*(t,S')$ implies $x \wedge x' \in X^*(t,S)$. Therefore,

$$\begin{aligned} f(x,t) \geq f(x \wedge x', t) &\implies x \in X^*(t,S) \\ &\implies x \vee x' \in X^*(t,S') \\ &\implies f(x \vee x', t) \geq f(x', t), \end{aligned}$$

and

$$\begin{aligned} f(x,t) > f(x \wedge x', t) &\implies x \wedge x' \notin X^*(t,S) \\ &\implies x' \notin X^*(t,S') \\ &\implies f(x \vee x', t) > f(x', t). \end{aligned}$$

Thus f is quasisupermodular in x. □

2.2.4 The General Theorem

Combining Theorems 2.13, 2.14, and 2.15, we obtain the following result from Milgrom and Shannon (1994, Theorem 4).

Theorem 2.16 (Milgrom and Shannon (1994)). Let X be a lattice, T be a partially ordered set, and $f : X \times T \to \mathbb{R}$. Then $\text{argmax}_{x \in S} f(x,t)$ is monotone nondecreasing in t and S (in the strong set order) if and only if f is quasisupermodular in x and has the single crossing property in $(x;t)$.

Proof. The "if" part of the theorem follows directly from Theorem 2.13.

To establish the "only if" part, suppose $\operatorname{argmax}_{x\in S} f(x,t)$ is monotone nondecreasing in t and S. To show that f satisfies the single crossing property in $(x;t)$, recall first that if $S=\{\bar{x},\bar{x}'\}$ for $\bar{x}'>\bar{x}$, then $S \geq_s S$. Thus, for any such set S, if $t' \geq t$, then

$$\operatorname*{argmax}_{x\in S} f(x,t') \geq_s \operatorname*{argmax}_{x\in S} f(x,t).$$

By Theorem 2.14, this implies that f satisfies the single crossing property in $(x;t)$. Finally, Theorem 2.15 implies that f is quasisupermodular in x. $\square$

2.3 Greatest and Least Solutions

When $X \subseteq \mathbb{R}$, the existence of a largest and smallest maximizer of an objective function f can be ensured by imposing standard topological assumptions, such as compactness of X and continuity of f. However, when $X \subseteq \mathbb{R}^n$, these assumptions are no longer sufficient. The difficultly is that a solution set $X^*(t)$ may not contain an element $\bar{x}$ that is greater than every other solution in *all* dimensions. In this section, we establish the existence of greatest and least maximizers under the additional assumptions that X is a lattice and f is a quasisupermodular function. We will use these results in the next section and then again in the following chapter when we study supermodular games.

We first establish three lemmas that will be combined together to yield our desired result about greatest and least solutions.

Lemma 2.17. Suppose X is a lattice, T is a partially ordered set, and $f: X\times T \to \mathbb{R}$. If f is quasisupermodular in x, then the solution set $X^*(t) = \operatorname{argmax}_{x\in X} f(x,t)$ is a sublattice of X for each t.

Proof. Fix any $x,x' \in X^*(t)$. Since X is a lattice, we have $x\wedge x', x\vee x' \in X$. To show that $x \vee x' \in X^*(t)$, note that

$$\begin{aligned} x \in X^*(t) &\implies f(x,t) \geq f(x\wedge x',t) \\ &\implies f(x\vee x',t) \geq f(x',t) \qquad \text{(QSM)} \\ &\implies x\vee x' \in X^*(t). \end{aligned}$$

Next, we prove that $x \wedge x' \in X^*(t)$ by contradiction:

$$\begin{aligned} x \wedge x' \notin X^*(t) &\implies f(x,t) > f(x \wedge x', t) \\ &\implies f(x \vee x', t) > f(x', t) \qquad \text{(QSM)} \\ &\implies x' \notin X^*(t), \end{aligned}$$

a contradiction to the assumption that $x' \in X^*(t)$. Therefore, it must be the case that $x \wedge x' \in X^*(t)$. □

If this proof seemed somewhat familiar, there is a good reason why. The proof of Theorem 2.12 used similar arguments. In fact, while a direct proof of this lemma is instructive, it is unnecessary: Applying Theorem 2.12 to the trivial parameter set $T = \{t\}$ implies that $X^*(t) \geq_s X^*(t)$. Then, recall that a set dominates itself in the strong set order if and only if it is a sublattice. Thus, $X^*(t)$ is a sublattice for each t.

Lemma 2.18. Suppose $X \subseteq \mathbb{R}^n$ is a nonempty lattice and is compact.[4] Then, X has a greatest and least element, that is, there exist $\underline{x}, \bar{x} \in X$ such that $\underline{x} \leq x \leq \bar{x}$ for all $x \in X$.

Proof. We will prove the existence of a greatest element $\bar{x} \in X$. The existence of a least element follows from a similar argument. Since X is compact, the set $\operatorname{argmax}_{x \in X} x_i$ (the set of maximizers of the continuous function $g(x) = x_i$) is nonempty. For each $i \in \{1, \dots, n\}$, fix $\hat{x}^i \in \operatorname{argmax}_{x \in X} x_i$. Then $x_i \leq \hat{x}^i_i$ for all $x \in X$. Let $\bar{x} = \hat{x}^1 \vee \hat{x}^2 \vee \cdots \vee \hat{x}^n = (\hat{x}^1_1, \dots, \hat{x}^n_n)$. That is, $\bar{x}$ is the coordinate-wise maximum of all of the vectors $\hat{x}^i$. In particular, for any $x \in X$ and any $i \in \{1, \dots, n\}$, $x_i \leq \hat{x}^i_i = \bar{x}_i$. It remains only to show that $\bar{x} \in X$. Since X is a lattice and $\hat{x}^1, \hat{x}^2 \in X$, we have $\hat{x}^1 \vee \hat{x}^2 \in X$. Repeating this argument, we also have $\hat{x}^1 \vee \hat{x}^2 \vee \hat{x}^3 = (\hat{x}^1 \vee \hat{x}^2) \vee \hat{x}^3 \in X$. Continuing in this manner, it follows by induction that $\bar{x} = \hat{x}^1 \vee \hat{x}^2 \vee \cdots \vee \hat{x}^n \in X$. □

Lemma 2.19. Suppose $Y, Z \subseteq \mathbb{R}^n$, and suppose each of these sets has greatest and least elements, $\bar{y}, \underline{y} \in Y$ and $\bar{z}, \underline{z} \in Z$, respectively. If $Z \geq_s Y$ then $\bar{z} \geq \bar{y}$ and $\underline{z} \geq \underline{y}$.

You are asked to prove this lemma in Exercise 2.10. Note that Lemma 2.19 extends the one-dimensional monotonicity result from Lemma 1.3 to multi-dimensional Euclidean spaces.

Theorem 2.20. Suppose X is a nonempty lattice and is compact, T is a partially ordered set, and $f : X \times T \to \mathbb{R}$. If f is continuous in x and quasisupermodular in x, then:

(1) The solution set $X^*(t) = \operatorname{argmax}_{x \in X} f(x,t)$ is nonempty and has greatest and least elements $\bar{x}(t)$ and $\underline{x}(t)$, respectively, for each $t \in T$.

[4] This is a special case of what is referred to as a **complete lattice**.

(2) If f also has the single crossing property in $(x;t)$, then $t' \geq t$ implies $\bar{x}(t') \geq \bar{x}(t)$ and $\underline{x}(t') \geq \underline{x}(t)$.

Proof. Since X is compact and f is continuous in x, the usual arguments imply that $X^*(t)$ is nonempty and compact for each t.[5] Since f is quasisupermodular in x, Lemma 2.17 implies $X^*(t)$ is also a sublattice of X. Then, since $X^*(t)$ is a lattice and is compact, Lemma 2.18 implies there exist $\underline{x}(t), \bar{x}(t) \in X^*(t)$ such that $\underline{x}(t) \leq x \leq \bar{x}(t)$ for all $x \in X^*(t)$.

If f also satisfies the single crossing property in $(x;t)$, then the set of maximizers $X^*(t)$ is monotone nondecreasing in t in the strong set order by Theorem 2.16. That is, $t' \geq t$ implies $X^*(t') \geq_s X^*(t)$. By Lemma 2.19, this implies $\bar{x}(t') \geq \bar{x}(t)$ and $\underline{x}(t') \geq \underline{x}(t)$. □

2.4 Application: Le Chatelier Principle

It is a well-known principle in economics that long-run demand is typically more elastic than short-run demand. For example, a change in the price of labor may lead to a change in the use of labor in the short run, even if capital is temporarily fixed. In the long run, capital may also change, thereby altering the marginal product of labor. This feedback leads to an additional change in labor use.

More concretely, depending or whether capital or labor are complements or substitutes, standard economic intuition would suggest one of two possible scenarios for the response of labor to a decrease in the wage rate:

- *Complements:* Labor increases in the short run ⇒ capital increases in the long run ⇒ labor increases even more in the long run.
- *Substitutes:* Labor increases in the short run ⇒ capital decreases in the long run ⇒ labor increases even more in the long run.

Notice that in both cases, the long-run response of labor is greater than the short-run response. Paul Samuelson coined the term "Le Chatelier Principle" to refer this property after the related principle in Physics.

The intuition provided above is very informal and is in fact incorrect when arbitrary production functions are allowed. Consider the following stylized example taken from Milgrom and Roberts (1996).

Example 2.21 (Labor response smaller in the long run). Let the feasible triples of labor, capital, and output (l, k, q) be the convex hull of the following set of three points:

[5] Fix any t, and treat f as a function of only x. The existence of a maximizer follows from the Weierstrass maximum theorem. If we let $v = \max_{x \in X} f(x,t)$, then the solution set $X^*(t) = \operatorname{argmax}_{x \in X} f(x,t)$ is the inverse image of this maximum value, $X^*(t) = \{x \in X : f(x,t) = v\}$. Since the inverse image of a closed set under a continuous function is closed (in fact, this is one way of defining continuity), $X^*(t)$ is closed. As a closed subset of the compact set X, the set $X^*(t)$ is therefore compact.

$\{(0,0,0),(2,0,1),(1,1,1)\}$. If the price of output is 2 and the initial wage and capital rental rates satisfy $w < r < 1$, then the initial optimal input mix is $(l,k) = (2,0)$. If the wage rises to some value $1 < w < 2-r$, then the new short-run optimal input mix is $(0,0)$, but the new long-run optimal is $(1,1)$. The demand for labor falls more in the short run than in the long run.

This example shows that not every production technology generates higher demand elasticity in the long run. However, under certain restrictions on the production technology, the conclusion of greater long-run elasticity of demand is valid. Extending Paul Samuelson's early formal work on the subject, Milgrom and Roberts (1996) applied the methods of monotone comparative statics to the problem and demonstrated that the Le Chatelier Principle holds for both complements and substitutes.

We begin with a general objective function and subsequently specialize the result to the input demand decision of a firm. In the abstract statement of the problem, there are two choice variables, $x \in X$ and $y \in Y$, where $X \subseteq \mathbb{R}$ and $Y \subseteq \mathbb{R}$.[6] The objective function $f : X \times Y \times T \to \mathbb{R}$ depends on a parameter t. To simplify exposition, in the case of multiple solutions our analysis will focus on the greatest maximizers. Thus, for each parameter t, let $(x^*(t), y^*(t))$ denote the greatest *long-run* solution. That is,

$$(x^*(t), y^*(t)) \in \operatorname*{argmax}_{(x,y)\in X\times Y} f(x,y,t)$$

and

$$(x^*(t), y^*(t)) \geq (x', y') \quad \forall (x', y') \in \operatorname*{argmax}_{(x,y)\in X\times Y} f(x,y,t).$$

From Theorem 2.20, we know that this greatest solution exists if X and Y are compact and f is continuous and quasisupermodular in (x, y). Moreover, if f has the single crossing property in $(x, y; t)$, then $x^*(t)$ and $y^*(t)$ are nondecreasing in t.

While both variables can be freely adjusted in the long run, suppose that only x can be varied in the short run. We are therefore interested in the optimal x conditional on both a parameter t and the temporarily-fixed value y. For any fixed y and t, let $x^s(y, t)$ be the largest *short-run* solution. That is,

$$x^s(y,t) \in \operatorname*{argmax}_{x\in X} f(x,y,t)$$

and

$$x^s(y,t) \geq x' \quad \forall x' \in \operatorname*{argmax}_{x\in X} f(x,y,t).$$

After a change in the parameter from t to t', y remains fixed at $y^*(t)$ in the short run, and hence the optimal choice of x is temporarily $x^s(y^*(t), t')$. However, in the long run,

[6]The results in this section immediately generalize to the case where X and Y are lattices in a multidimensional space.

x and y are both changed to their optimal long-run values $x^*(t')$ and $y^*(t')$. Note that $x^*(t') = x^s(y^*(t'), t')$ since the short- and long-run solutions for x are the same when y takes its optimal long-run value.

Theorem 2.22 (Milgrom and Roberts (1996)). Suppose $X \subseteq \mathbb{R}$ and $Y \subseteq \mathbb{R}$ are compact, T is a partially ordered set, and $f : X \times Y \times T \to \mathbb{R}$. If f is continuous and quasisupermodular in (x, y) and has the single crossing property in $(x, y; t)$, then for any $t' \geq t$,

$$\begin{aligned} x^*(t) &\leq x^s(y^*(t), t') \leq x^*(t'), \quad \text{and} \\ x^*(t) &\leq x^s(y^*(t'), t) \leq x^*(t'). \end{aligned}$$

Proof. As noted above, Theorem 2.20 implies that $x^*(t)$ and $y^*(t)$ are nondecreasing in t. Thus $y^*(t') \geq y^*(t)$. Also note that $x^s(y, t)$ is nondecreasing in t for fixed y, which follows by applying Theorem 2.20 with x treated as the variable, t as the parameter, and y fixed as a constant;[7] it is also nondecreasing in y for fixed t by the same theorem since quasisupermodularity of f in (x, y) implies the single crossing property in $(x; y)$ (see Exercise 2.6). Changing one variable at a time, we therefore obtain:

$$\begin{aligned} x^s(y^*(t), t) &\leq x^s(y^*(t), t') \leq x^s(y^*(t'), t'), \quad \text{and} \\ x^s(y^*(t), t) &\leq x^s(y^*(t'), t) \leq x^s(y^*(t'), t'). \end{aligned}$$

Since $x^*(t) = x^s(y^*(t), t)$ for any t, this completes the proof. □

The first set of inequalities displayed in Theorem 2.22 captures the short-run and long-run response to an increase in the parameter from t to t'. Note that y is fixed at $y^*(t)$ in the short run. The second set of inequalities captures the response to a decrease in the parameter from t' to t. In this case y is fixed at $y^*(t')$ in the short run. For both directions of a change in the parameter, the variable x responds more in the long run than in the short run.

One particular application of interest is a firm's optimal labor demand when capital is fixed in the short run. Letting g denote the production function, the profit function is

$$\pi(l, k, w) = pg(l, k) - wl - rk.$$

We are not including the output price p and rental rate of capital r as arguments in this function since we are only considering variation in the wage in what follows. Define the largest solutions for long-run labor and capital, $l^*(w)$ and $k^*(w)$, and the largest solution for short-run labor, $l^s(k, w)$, just as in the case of general objective functions above.

[7] Here, we could also simply apply our single-dimensional single crossing result (Theorem 2.2) to show that the set of maximizing x are nondecreasing in t (in the strong set order) when y is held fixed. We would then apply Corollary 1.5 to conclude that there exists a greatest maximizing x and it is nondecreasing in t.

Corollary 2.23. Suppose $L \subseteq \mathbb{R}$ and $K \subseteq \mathbb{R}$ are compact and $g : L \times K \to \mathbb{R}$ is continuous. If g is either supermodular or submodular, then for any $w > w'$,[8]

$$\begin{aligned} l^*(w) &\leq l^s(k^*(w), w') \leq l^*(w') \\ l^*(w) &\leq l^s(k^*(w'), w) \leq l^*(w'). \end{aligned}$$

When g is supermodular, labor and capital are complements; when g is submodular, they are instead substitutes. Note that in either case, Theorem 2.22 can be applied after an appropriate transformation of the variables. You are asked to complete the details in Exercise 2.11.

2.5 Exercises

2.1 Answer the following. In what follows, do *not* assume that any of the functions are differentiable.

(a) Suppose $X \subseteq \mathbb{R}$ and $T \subseteq \mathbb{R}$, and suppose that $f : X \times T \to \mathbb{R}$ and $g : X \times T \to \mathbb{R}$ have increasing differences in $(x;t)$. Does the function $h : X \times T \to \mathbb{R}$ defined by $h(x,t) = f(x,t) + g(x,t)$ have increasing differences in $(x;t)$? Prove or provide a counterexample.

(b) Suppose $X \subseteq \mathbb{R}$ and $T \subseteq \mathbb{R}$, and suppose that $f : X \times T \to \mathbb{R}$ and $g : X \times T \to \mathbb{R}$ have the single crossing property in $(x;t)$. Does the function $h : X \times T \to \mathbb{R}$ defined by $h(x,t) = f(x,t) + g(x,t)$ have the single crossing property in $(x;t)$? Prove or provide a counterexample.

2.2 This problem concerns the strict single crossing property.

(a) Prove Theorem 2.8.

(b) Prove Corollary 2.9.

(c) If we change the statement of Corollary 2.9 by replacing "strict single crossing" with "single crossing", is it still true? Prove or provide a counterexample.

(d) If we change the statement of Corollary 2.9 by replacing "nondecreasing in t" with "strictly increasing in t", is it still true? Prove or provide a counterexample.

2.3 Suppose $X = \mathbb{R}_+$ and $T = \mathbb{R}^2_+$. Let $f(x,t) = xg(t_1 + t_2) - x^2$ for some function $g : \mathbb{R}_+ \to \mathbb{R}_+$.

(a) Find a condition on g that is necessary and sufficient for f to have increasing differences in $(x;t)$.

[8] In the case where g is submodular, let $l^*(w)$ again be the greatest solution for long-run labor, but let $k^*(w)$ instead be the smallest solution for long-run capital. This change is important for technical reasons, since a change of variables will be required to prove the result for the case of submodular g; however, it obviously does not effect the economic interpretation of the result.

(b) Find a condition on g that is necessary and sufficient for f to be supermodular in t.

(c) Find a condition on g that is necessary and sufficient for $\operatorname{argmax}_{x \in \mathbb{R}_+} f(x,t)$ to be monotone nondecreasing in t.

(d) Is supermodularity or quasisupermodularity of f in t necessary for MCS? Explain why or why not.

2.4 Suppose $X \subseteq \mathbb{R}^n$ is a lattice, T is a partially ordered set, $f : X \times T \to \mathbb{R}$, and $g : \mathbb{R} \times T \to \mathbb{R}$ is strictly increasing in its first variable for every $t \in T$.

(a) Prove that if $g(f(x,t),t)$ has increasing differences in $(x;t)$ then f satisfies the single crossing property in $(x;t)$.

(b) Prove that if $g(f(x,t),t)$ is supermodular in x then f is quasisupermodular in x.

2.5 Suppose $X \subseteq \mathbb{R}$, $T \subseteq \mathbb{R}$ and $f : X \times T \to \mathbb{R}_{++}$, that is, $f(x,t) > 0$ for every pair $(x,t) \in X \times T$. The following diagram illustrates the logical relationships between various properties of f and $\log f$:

$$\begin{array}{ccc} f \text{ has ID in } (x;t) & \xRightarrow{(1)} & f \text{ has SCP in } (x;t) \\ (4)\ \not\Downarrow \quad \not\Uparrow\ (5) & & (3)\ \Updownarrow \\ \log f \text{ has ID in } (x;t) & \xRightarrow{(2)} & \log f \text{ has SCP in } (x;t) \end{array}$$

Prove each of the relationships (1)–(5) illustrated above. (Note that (4) and (5) are negative relationships, meaning that you should show that they do not hold in general.)

2.6 Suppose $X_1 \subseteq \mathbb{R}$ and $X_2 \subseteq \mathbb{R}$. Let $X = X_1 \times X_2 \subseteq \mathbb{R}^2$ (note that X is therefore a lattice). Suppose $f : X \times T \to \mathbb{R}$, where T is a partially ordered set. We will write $x = (x_1, x_2)$ for elements of X and t for elements of T. Do *not* assume that f is differentiable.

(a) Prove that f is supermodular in x if and only if f has increasing differences in $(x_1; x_2)$ (for every fixed t).

(b) Prove that if f is quasisupermodular in x, then f satisfies the single crossing property in $(x_1; x_2)$ (for every fixed t). (Optional: You can also show conversely that if f has the single crossing property in both $(x_1; x_2)$ and $(x_2; x_1)$, then f is quasisupermodular in x.)

(c) Suppose f has increasing differences in $(x_1; t)$ (for every fixed x_2) and in $(x_2; t)$ (for every fixed x_1). Does this imply that f has increasing differences in $(x; t)$? Prove or provide a counterexample.

(d) Suppose f has the single crossing property in $(x_1; t)$ (for every fixed x_2) and in $(x_2; t)$ (for every fixed x_1). Does this imply that f has the single crossing property in $(x; t)$? Prove or provide a counterexample.

2.7 Suppose $X \subseteq \mathbb{R}^2$ is a lattice, and suppose that $f : X \to \mathbb{R}$ is defined by $f(x) = g(x_1 + x_2)$ for $x = (x_1, x_2) \in X$, where $g : \mathbb{R} \to \mathbb{R}$. Suppose there exists some $\alpha \in \mathbb{R}$ such that g is strictly decreasing up to α and strictly increasing beyond α; that is, $y < y' \leq \alpha$ implies $g(y) > g(y')$, and $\alpha \leq y < y'$ implies $g(y) < g(y')$. Is f quasisupermodular in x? Prove or provide a counterexample.

2.8 Suppose a firm has production function $f(x)$ for inputs $x \in \mathbb{R}^n_+$. Assume that f is nondecreasing but make no other assumptions. Both the input and the output markets are competitive: The price of output is $p > 0$ and the price of inputs is given by the vector $w \in \mathbb{R}^n_{++}$.

(a) Starting with a concrete example, suppose $n = 2$ and $f(x) = x_1^\alpha x_2^\beta$ for $\alpha, \beta \in (0, 1)$ with $\alpha + \beta < 1$. Solve for optimal inputs as a function of the prices p, w_1, and w_2. What is the relationship between w_i and x_i, and what is the relationship between w_i and x_j for $j \neq i$?

(b) Move back to the general case of any n and an unspecified production function f. Is it necessarily the case that $X_i^*(p, w)$ is nonincreasing in w_i? If yes, prove it. If not, briefly explain why not and give the weakest condition that you can find on f that is sufficient for $X_i^*(p, w)$ to be nonincreasing in w_i. Prove that your condition is sufficient.

(c) Is it necessarily the case that $X_i^*(p, w)$ is nonincreasing in w_j for $j \neq i$? If yes, prove it. If not, briefly explain why not and give the weakest condition that you can find on f that is sufficient for $X_i^*(p, w)$ to be nonincreasing in w_j. Prove that your condition is sufficient.

(d) Suppose $n = 2$. Is it necessarily the case that $X_i^*(p, w)$ is *nondecreasing* in w_j for $j \neq i$? If yes, prove it. If not, briefly explain why not and give the weakest condition that you can find on f that is sufficient for $X_i^*(p, w)$ to be nondecreasing in w_j. Prove that your condition is sufficient.

(e) Provide an example of a production function on $\mathbb{R}^2$ that has the properties you found in part (d).

2.9 A consumer has preferences over amounts of money $m \in \mathbb{R}$ and amounts of consumption of $n \geq 1$ non-money goods $x \in \mathbb{R}^n_+$. Assume that her preferences are continuous, strictly monotonic, and quasilinear in money. The utility of bundle (m, x) is

$$U(m, x) = m + v(x)$$

where $v : \mathbb{R}^n_+ \to \mathbb{R}$ is twice continuously differentiable. The consumer has wealth w. Bundle (m, x) costs $m + p \cdot x$, where $p \in \mathbb{R}^n_{++}$ is the non-money goods price vector and the price of money is fixed at 1. Note that although each x_i must take nonnegative values, m is permitted to take negative values (this is a standard assumption when working with quasilinear utility). Let $M^*(p)$ and $X^*(p)$ denote the demands for money and non-money goods, respectively.

For the first parts of the problem, assume that $n = 1$:

(a) Is it necessarily the case that $X^*(p)$ is nonincreasing in p? If yes, prove it. If not, briefly explain why not, and give the weakest condition that you can find on v that is sufficient for $X^*(p)$ to be nonincreasing in p. Prove that your condition is sufficient.

(b) Find a condition in terms of $v''(\cdot)$ and $v'(\cdot)$ under which $M^*(p)$ is nondecreasing in p.

(c) Either using your results from the previous part of the problem or by direct calculation, prove that for $v(x) = \ln(x)$ we have that $M^*(p)$ is constant in p.

(d) Argue that there are functions for which $M^*(p)$ is strictly increasing in p and functions for which $M^*(p)$ is strictly decreasing in p. (Hint: Although we have not discussed general theorems giving conditions for strictly increasing or decreasing solutions, in this case you are just looking for utility functions that give closed-formed solutions with these properties. If you solved part (b) correctly, then it may give you a hint of a parametric family of functions that you could use and that includes $v(x) = \ln(x)$ as a special case.)

Now assume that $n > 1$:

(e) Is it necessarily the case that $X^*(p)$ is nonincreasing in p? If yes, prove it. If not, briefly explain why not, and give the weakest condition that you can find on v that is sufficient for $X^*(p)$ to be nonincreasing. Prove that your condition is sufficient.

(f) Assume any conditions that you found in part (e). Let $V(p)$ denote the consumer's indirect utility function (as a function of p only). Prove that V is supermodular:

   i. Assuming $X^*(p)$ is singleton-valued, and hence has a unique selection function $x^*(p)$.

   ii. Without assuming $X^*(p)$ is singleton-valued.

2.10 Prove Lemma 2.19.

2.11 This problem continues the analysis of the Le Chatelier Principle for a firm's labor decision.

(a) Prove Corollary 2.23. Your proof should include two parts, one for each of the two cases considered in the corollary (complements and substitutes). Be specific about how the optimal long-run capital $k^*(w)$ relates to $k^*(w')$ in each of these two cases.

(b) Can the assumption of supermodularity (submodularity) of g be replaced with quasisupermodularity (quasisubmodularity) in Corollary 2.23? Explain why or why not.

2.12 Suppose a firm has the following constant elasticity of substitution (CES) production function:

$$f(x_1, x_2) = (\alpha x_1^\gamma + (1-\alpha)x_2^\gamma)^{\beta/\gamma},$$

for $\gamma \in \mathbb{R}$ (and $\gamma \neq 0$), $\alpha \in (0,1)$, and $\beta > 0$.[9] Assume that both the input and output markets are competitive, so the firm is a price taker. The market price of output is $p > 0$, and the price of inputs are $w_1, w_2 > 0$. Taking these prices as given, the firm chooses $x_1, x_2 \in [0, \bar{x}]$ to maximize profit. You can assume throughout that if there are multiple solutions, the firm chooses the solution with the highest value of x_2 (and if there are multiple solutions with this value of x_2, the firm chooses the one with the highest value of x_1 conditional on this x_2).

(a) Under what restrictions on the parameter values are the inputs complements (meaning an increase in the price of one input weakly decreases the optimal quantity of the other)?

(b) Suppose x_2 is fixed in the short run (e.g, capital), and x_1 is variable in both the long run and the short run (e.g., labor). Under the parameter restrictions you found in part (a), is the change in x_1 in response to an increase in w_1 smaller in the long run than in the short run, larger in the long run than in the short run, or is the comparison of the magnitude of these two changes indeterminate under these assumptions? If you draw a definitive conclusion about which response is larger, provide precise arguments in support of your conclusion. If you claim there is insufficient information to determine whether the change is larger in the short run or the long run, provide a careful explanation of why this is the case (but counterexamples are not required for this part).

(c) Under what restrictions on the parameter values are the inputs substitutes (meaning an increase in the price of one input weakly increases the optimal quantity of the other)?

(d) Suppose x_2 is fixed in the short run (e.g, capital), and x_1 is variable in both the long run and the short run (e.g., labor). Under the parameter restrictions you found in part (c), is the change in x_1 in response to an increase in w_1 smaller in the long run than in the short run, larger in the long run than in the short run, or is the comparison of the magnitude of these two changes indeterminate under these assumptions? If you draw a definitive conclusion about which response is larger, provide precise arguments in support of your conclusion. If you claim there is insufficient information to determine whether the change is larger in the short run or the long run, provide a careful explanation of why this is the case (but counterexamples are not required for this part).

[9] This production function is called CES because the elasticity of substitution between the inputs is $\sigma = 1/(1-\gamma)$.

Chapter 3

Supermodular Games

Contents

References and Assigned Readings

Primary readings:

(1) Milgrom and Roberts (1990) (recommended).
(2) Levin notes on SM Games from his game theory course (optional).

Additional references:

(1) Muhamet Yildiz course notes on Supermodular Games.
(2) Topkis (1998)
(3) Vives (1990)

3.1 Introduction and Definitions

3.1.1 Examples of Strategic Complements and Substitutes

Before delving into the formal definition of a supermodular game, we begin with a few simple motivating examples that illustrate when the strategies of different players are complementary in the sense that one player increasing zir strategy leads the other player to want to increase zir strategy in response. Such games are said to have **strategic complements**.

EXAMPLE 3.1 (PRICE COMPETITION WITH DIFFERENTIATED PRODUCTS).

$$\begin{aligned} &D_i(p_i, p_{-i}) = a_i - b_i p_i + \sum_{j \neq i} d_{ij} p_j, \qquad b_i, d_{ij} \geq 0 \\ &\pi_i(p_i, p_{-i}) = (p_i - c_i) D_i(p_i, p_{-i}) \\ &\frac{\partial^2 \pi_i}{\partial p_i \partial p_j} = d_{ij} \geq 0. \end{aligned}$$

The profit function has increasing differences in $(p_i; p_{-i})$. The best response to an increase in p_{-i} is to increase p_i: This is a game with strategic complements.

EXAMPLE 3.2 (SEARCH). This is a simplified version of the Diamond (1982) search model. Suppose agents exert effort to look for trading partners.

$$u_i(e_i, e_{-i}) = e_i \cdot \sum_{j \neq i} e_j - c(e_i).$$

This utility function has increasing differences in $(e_i; e_{-i})$. Hence, this is also a game of strategic complements.

There are other situations where the strategies of the players are substitutes in the sense that one player increasing zir strategy leads the other player to want to decrease zir strategy. Such games are said to have **strategic substitutes**. The following is an example of such a game.

Example 3.3 (Cournot Duopoly).

$$\begin{aligned} &P(q_1, q_2) = a - b(q_1 + q_2), \qquad b \geq 0 \\ &\pi_i(q_i, q_j) = (P(q_1, q_2) - c_i)q_i \\ &\frac{\partial^2 \pi_i}{\partial q_i \partial q_j} = -b \leq 0. \end{aligned}$$

Thus, π_i has increasing differences in $(q_i; -q_j)$. The best response to an increase in q_j is to decrease q_i: This is a game with strategic substitutes.

The first two examples are special cases of what we will refer to as supermodular games. The third example is not, but we will see that after a change of variables, it can be transformed into a supermodular game. We next define these concepts formally and then explore some of the properties of supermodular games.

3.1.2 Supermodular Games

Consider an n-player game. Each player has a strategy space S_i that is a subset of $\mathbb{R}^{m_i}$ for some positive integer $m_i \in \mathbb{N}$. Thus we can write $s_i = (s_i^1, \ldots, s_i^{m_i}) \in S_i$ to denote a pure strategy for player i. We denote the complete strategy profile by

$$\begin{aligned} s &= (s_1, \ldots, s_n) \\ &= (s_1^1, \ldots, s_1^{m_1}, s_2^1, \ldots, s_2^{m_2}, \ldots, s_n^1, \ldots, s_n^{m_n}) \in S_1 \times \cdots \times S_n \subseteq \mathbb{R}^m, \end{aligned}$$

where $m = m_1 + \cdots + m_n$. The generality of multi-dimensional strategy spaces will be important for some of our applications, but for intuition you can simply think of each S_i as a subset of $\mathbb{R}$ and think of S as a subset of $\mathbb{R}^n$.

Let

$$S \equiv \prod_{i \in N} S_i = S_1 \times \cdots \times S_n \qquad \text{and} \qquad S_{-i} \equiv \prod_{j \neq i} S_j.$$

Following conventional game theory notation, we will often write $(s_i, s_{-i}) \in S$ to denote the strategy profile obtained when player i uses strategy $s_i \in S_i$ and the profile of strategies of the other players is $s_{-i} \in S_{-i}$.

Definition 3.4. Let $N = \{1, \ldots, n\}$ denote the set of players. A normal-form game $(N, (S_i)_{i \in N}, (u_i)_{i \in N})$ is a **supermodular game** if for each $i \in N$:

(1) $S_i \subseteq \mathbb{R}^{m_i}$ is a lattice and is compact.[1]
(2) $u_i(s_i, s_{-i})$ is continuous in s_i for fixed s_{-i}.[2]
(3) $u_i(s_i, s_{-i})$ is quasisupermodular in s_i and satisfies the single crossing property in $(s_i; s_{-i})$.

The classic definition of a supermodular game due to Topkis (1979) assumes each player i has a utility function u_i that is supermodular in their own strategy s_i and has increasing differences in $(s_i; s_{-i})$. The important implication of these assumptions in the analysis of the game is that the set of best responses of each player is nondecreasing (in the strong set order) in the strategies of the other players; in particular, the greatest and least best responses are nondecreasing. We have therefore used the more general ordinal notions of quasisupermodularity and the single crossing property from Milgrom and Shannon (1994) in Definition 3.4, but have retained the original name.

The following theorem proves the existence and monotonicity of the greatest and least best response functions for each player in a supermodular game. Define the **best-response correspondence** $B_i : S_{-i} \twoheadrightarrow S_i$ for player i by

$$B_i(s_{-i}) = \operatorname*{argmax}_{s_i \in S_i} u_i(s_i, s_{-i}).$$

Theorem 3.5. Suppose $(N, (S_i)_{i\in N}, (u_i)_{i\in N})$ is a supermodular game. Then:

(1) $B_i(s_{-i})$ is nonempty and has greatest and least elements $\overline{B}_i(s_{-i})$ and $\underline{B}_i(s_{-i})$, respectively.
(2) If $s'_{-i} \geq s_{-i}$ then $\overline{B}_i(s'_{-i}) \geq \overline{B}_i(s_{-i})$ and $\underline{B}_i(s'_{-i}) \geq \underline{B}_i(s_{-i})$.

The set of best responses $B_i(s_{-i})$ of player i is the set of optimal strategies s_i, taking the strategy profile s_{-i} of the other players as given. Translating into the notation of the previous chapters, the objective function f is the utility function u_i, the choice variable x is the strategy s_i, and the parameter t is the strategy profile s_{-i} of the other players. Theorem 3.5 is therefore a special case of Theorem 2.20.

In the case where players have single-dimensional strategy spaces—which is again what you should be thinking of for intuition—the result simply says that there are upper and lower bounds in $B_i(s_{-i})$ and these are nondecreasing in s_{-i}. In the general case with multidimensional strategy spaces, $B_i(s_{-i})$ is a set of vectors $s_i \in \mathbb{R}^{m_i}$. Theorem 3.5 shows that there exist best responses that are greatest (and least) in all of the m_i coordinates.

[1] This is a special case of what is referred to as a **complete lattice**.

[2] Upper semi-continuity in s_i is in fact sufficient for our first set of results, since continuity will only be used to show the existence of a compact set of best responses to each profile of strategies of the other players.

3.2 Equilibrium Existence

For a given normal-form game, define the combined **best-response correspondence** of all players $B : S \twoheadrightarrow S$ by

$$B(s) = \prod_{i \in N} B_i(s_{-i}) = B_1(s_{-1}) \times \cdots \times B_n(s_{-n}).$$

Recall that the fixed points of this correspondence are precisely the pure-strategy Nash equilibria of the game:

$$s \in B(s) \iff s \text{ is a Nash equilibrium.}$$

In other courses, you have already seen one approach to proving the existence of a Nash equilibrium that relies on the Kakutani fixed point theorem and therefore assumes continuity and convexity properties. In particular, Kakutani's theorem can be applied if each S_i is nonempty, compact, and convex and if the correspondence B is upper hemi-continuous, nonempty-valued, and convex-valued, conditions that are satisfied if u_i is continuous in (s_i, s_{-i}) and quasi-concave in s_i.[3]

In this section, we show that these assumptions can be relaxed when B is monotonic, as in a supermodular game. These results provide a useful alternative approach to establishing the existence of a pure-strategy Nash equilibrium in games where continuity and quasi-concavity may be violated, or where the strategy space S_i is not convex. We will further show that the set of equilibria in a supermodular game has a useful structure.

For a given supermodular game, define functions $\overline{B} : S \to S$ and $\underline{B} : S \to S$ by

$$\begin{aligned}
\overline{B}(s) &= (\overline{B}_i(s_{-i}))_{i \in N} = (\overline{B}_1(s_{-1}), \ldots, \overline{B}_n(s_{-n})) \\
\underline{B}(s) &= (\underline{B}_i(s_{-i}))_{i \in N} = (\underline{B}_1(s_{-1}), \ldots, \underline{B}_n(s_{-n})).
\end{aligned}$$

Thus, $\overline{B}$ and $\underline{B}$ are greatest and least best-response *functions* that bound the best-response correspondence B. Note that these functions can be used to determine a set of sufficient (but not necessary) conditions for a strategy profile to be a Nash equilibrium:

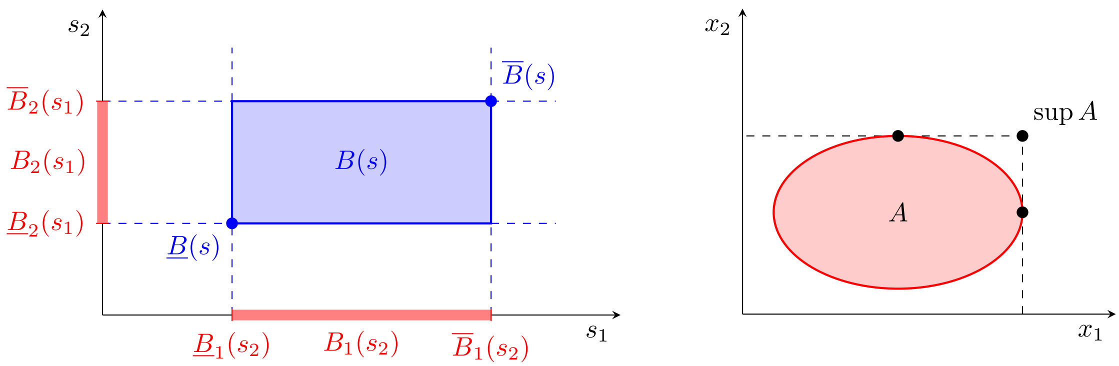
$$\begin{aligned}
s = \overline{B}(s) &\implies s \text{ is a Nash equilibrium} \\
s = \underline{B}(s) &\implies s \text{ is a Nash equilibrium.}
\end{aligned}$$

Importantly, by Theorem 3.5, both $\overline{B}$ and $\underline{B}$ are nondecreasing functions. This property will

[3] What about mixed strategies? If we assume expected-utility preferences over lotteries, then quasi-concavity of u_i in the lottery is immediate. If the game is finite, then continuity is also implied and the existence of a mixed-strategy Nash equilibrium therefore ensured. Unfortunately, existence results for finite games based on Kakutani's theorem do not guarantee the existence of a pure-strategy Nash equilibrium. It is also important to keep in mind that in *infinite* games, continuity of u_i in the lottery is not automatic—it requires continuity of u_i with respect to pure strategies. Thus, continuity is needed to prove even the existence of equilibrium in mixed strategies in infinite games using Kakutani's theorem.

be central to our approach to proving the existence of Nash equilibrium in supermodular games.

In the case of a two-player game in which each player's strategy space is a subset of the real line, Figure 3.1a illustrates the (two-dimensional) best-response set $B(s)$ and the greatest and least best responses $\overline{B}(s)$ and $\underline{B}(s)$ that arise from the best response sets $B_1(s_2)$ and $B_2(s_1)$ of the two players. While this figure depicts the best responses to a single (fixed) strategy profile $s = (s_1, s_2)$ (not pictured), we will of course be interested in how $\overline{B}(s)$ and $\underline{B}(s)$ change as s varies and where these functions have fixed points.



(a) Illustration of $B(s)$, $\overline{B}(s)$, and $\underline{B}(s)$

(b) Supremum of $A \subseteq \mathbb{R}^2$

Figure 3.1: Best response mappings and supremum of a two-dimensional set.

We will prove the existence of a Nash equilibrium by applying a fixed point theorem due to Tarski. Before stating this theorem, we need to formally define the supremum and infimum of subsets of a Euclidean space $\mathbb{R}^m$. These definitions extend the familiar definitions from the one-dimensional real line. For any nonempty and bounded set $A \subseteq \mathbb{R}^m$, the **supremum**, or least upper bound, of A is the vector $\bar{x} = \sup A$ that satisfies: (i) $x \leq \bar{x}$ for all $x \in A$, and (ii) if any other vector y satisfies $x \leq y$ for all $x \in A$ then $\bar{x} \leq y$. Note that if $A = \{x, x'\}$, then the supremum of A is simply the join of these two vectors: $\sup\{x, x'\} = x \vee x'$. In general, for any bounded set $A \subseteq \mathbb{R}^m$, we will use the notation $\sup A$ and $\bigvee A$ interchangeably. Figure 3.1b illustrates the supremum in the case of a set $A \subseteq \mathbb{R}^2$. We similarly write $\inf A$ or $\bigwedge A$ to denote the **infimum**, or greatest lower bound, of A.

Importantly, if $X \subseteq \mathbb{R}^m$ is a lattice and is compact, then for any nonempty subset $A \subseteq X$ the supremum and infimum of this set are also elements of X, that is, $\inf A, \sup A \in X$. It is a good exercise to verify this claim.[4] It is also a good exercise to check why the assumptions that X is a lattice and that it is compact are both needed.

Theorem 3.6 (Tarski fixed point theorem). Suppose $X \subseteq \mathbb{R}^m$ is a nonempty lattice that is compact, and suppose $f : X \to X$ is a nondecreasing function. Then f has a fixed

[4] Hint: The proof strategy will have a similar flavor to the proof of Lemma 2.18. First show that there exist vectors $\hat{x}^i$ in X such that the ith coordinate of $\hat{x}^i$ is the supremum of the ith coordinate of the elements of A. Then take the join $\bar{x} = \hat{x}^1 \vee \hat{x}^2 \vee \cdots \vee \hat{x}^m$.

point. Moreover,

$$\bar{x} \equiv \sup\{x \in X : x \leq f(x)\}$$

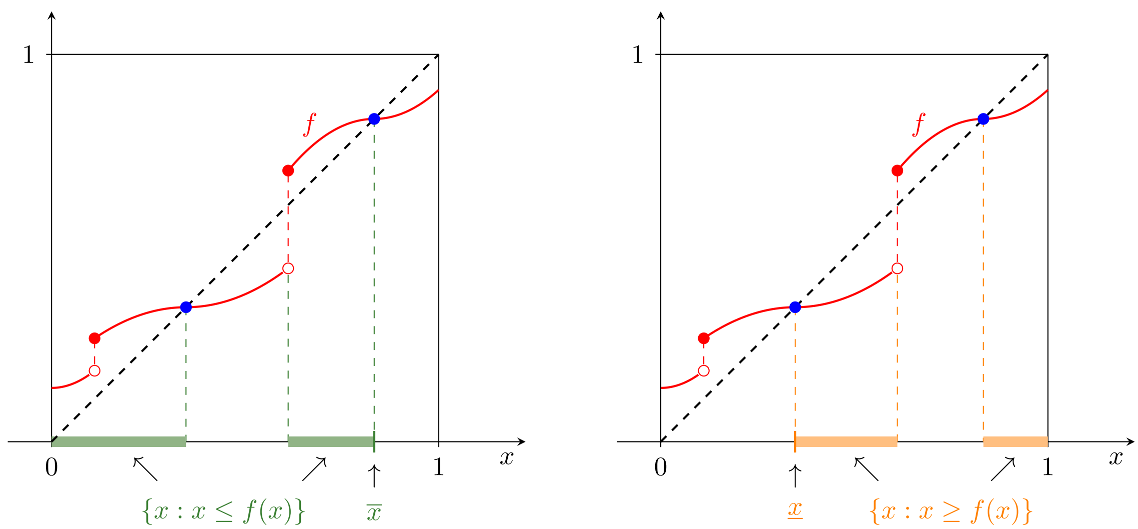
is the largest fixed point, and

$$\underline{x} \equiv \inf\{x \in X : f(x) \leq x\}$$

is the smallest fixed point. That is, if $x \in X$ and $f(x) = x$, then $\underline{x} \leq x \leq \bar{x}$.

Proof. Let $A = \{x \in X : x \leq f(x)\}$, and let $\bar{x} = \sup A$. (Note that A is nonempty since the least element of X must be contained in A.) By the definition of the supremum, for any $x \in A$, we have $x \leq \bar{x}$ and therefore $f(x) \leq f(\bar{x})$ since f is nondecreasing. Since $x \leq f(x)$ for all $x \in A$, we therefore have $x \leq f(x) \leq f(\bar{x})$ for all $x \in A$. Thus, $f(\bar{x})$ is an upper bound of the set A. Since $\bar{x}$ is the least upper bound, this implies $\bar{x} \leq f(\bar{x})$.

To show the opposite inequality, note that since $\bar{x} \leq f(\bar{x})$ and f is nondecreasing, we have $f(\bar{x}) \leq f(f(\bar{x}))$. That is, $f(\bar{x}) \in A$. Since $\bar{x}$ is an upper bound for A, this implies $f(\bar{x}) \leq \bar{x}$. Combining the two steps, we have shown that $f(\bar{x}) = \bar{x}$, so $\bar{x}$ is a fixed point. Moreover, since any fixed point of f must be in the set A, it follows that $\bar{x}$ is the largest fixed point. The proof that $\underline{x}$ is the smallest fixed point is similar. $\square$



(a) Largest fixed point

(b) Smallest fixed point

Figure 3.2: Illustration of the Tarski fixed point theorem.

Figure 3.2 illustrates the Tarski fixed point theorem, showing the two sets employed in the theorem and the largest and smallest fixed points. Note that unlike many other fixed point theorems used in economics, this result does not require that the function f be continuous. This assumption is replaced with the requirement that the function be nondecreasing. Also, despite the fact that the domain of f in this figure happens to be an interval, the set X in this theorem need not be convex; if fact, it could simply be a discrete set of points. However, it is essential for this result that the set X is a lattice, a requirement that is not needed for some other fixed point theorems.

Theorem 3.7 (Topkis (1979)). Suppose $(N, (S_i)_{i\in N}, (u_i)_{i\in N})$ is a supermodular game. Then a pure-strategy Nash equilibrium exists. Moreover,

$$\bar{s} = \sup\{s \in S : s \leq \overline{B}(s)\}$$

is the largest Nash equilibrium, and

$$\underline{s} = \inf\{s \in S : \underline{B}(s) \leq s\}$$

is the smallest Nash equilibrium. That is, if s is a pure-strategy Nash equilibrium, then $\underline{s} \leq s \leq \bar{s}$.

Proof. Define $\bar{s}$ as in the statement of the theorem. By Theorem 3.5, $\overline{B}(s)$ is nondecreasing in s. Therefore, $\bar{s}$ as defined above is a fixed point of $\overline{B}$ by Theorem 3.6 (i.e., $\bar{s} = \overline{B}(\bar{s})$), and hence $\bar{s}$ is a Nash equilibrium. To see that it is the largest Nash equilibrium, consider any other Nash equilibrium strategy profile $s \in S$. By definition, $s \in B(s)$ and therefore $s \leq \overline{B}(s)$. Thus $s \in \{s' \in S : s' \leq \overline{B}(s')\}$, which implies $s \leq \bar{s}$. The argument that $\underline{s}$ is the smallest Nash equilibrium is similar. $\square$

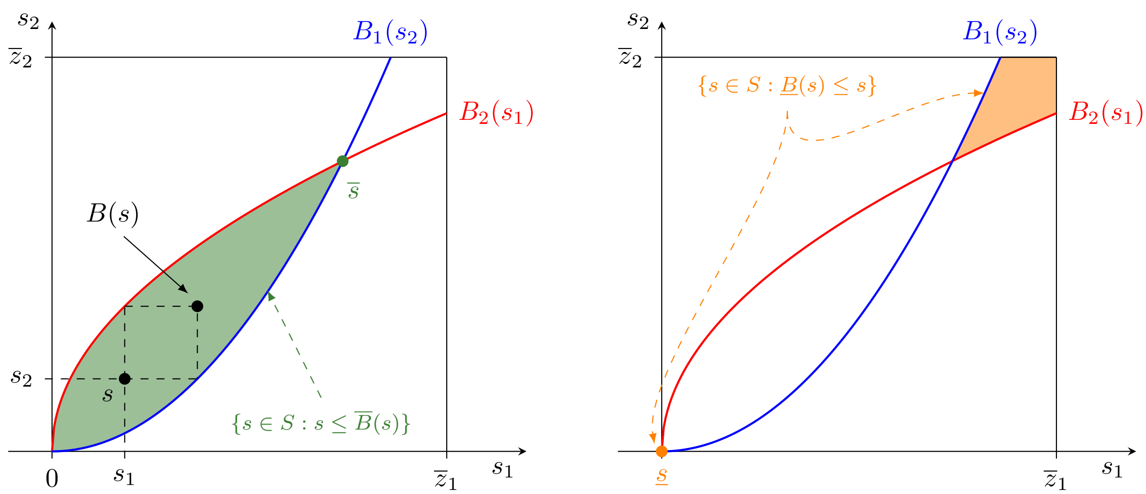
EXAMPLE 3.8 (PARTNERSHIP GAME). Suppose players 1 and 2 are engaged in a partnership relationship. Each player supplies an input $s_i \in [0, \bar{z}_i]$, and the output of the firm is $f(s_1, s_2) = s_1^\alpha s_2^\beta$, where $\alpha, \beta \in (0,1)$ and $\alpha + \beta < 1$. For example, s_1 could be capital and s_2 could be labor, or each s_i could be a different (and complementary) type of skilled labor. Note that f has increasing differences in $(s_1; s_2)$ (or equivalently, ID in $(s_2; s_1)$). The partners share output equally, and both find it costly to supply their input. Specifically, their utility functions are

$$u_1(s_1, s_2) = \frac{s_1^\alpha s_2^\beta}{2} - s_1 \quad \text{and} \quad u_2(s_1, s_2) = \frac{s_1^\alpha s_2^\beta}{2} - s_2.$$

For each strategy of the other player, there is a unique best response: $B_i(s_{-i}) = \{\underline{B}_i(s_{-i})\} = \{\overline{B}_i(s_{-i})\}$ where

$$\underline{B}_1(s_2) = \overline{B}_1(s_2) = \min\left\{\left(\frac{\alpha}{2}\right)^{1/(1-\alpha)} s_2^{\beta/(1-\alpha)},\ \bar{z}_1\right\}$$
$$\underline{B}_2(s_1) = \overline{B}_2(s_1) = \min\left\{\left(\frac{\beta}{2}\right)^{1/(1-\beta)} s_1^{\alpha/(1-\beta)},\ \bar{z}_2\right\}.$$

Note that $s \leq \overline{B}(s)$ is equivalent to $s_2 \leq \overline{B}_2(s_1)$ and $s_1 \leq \overline{B}_1(s_2)$, that is, each player's strategy is below her best response to the other player's strategy. The set of strategy profiles where $s \leq \overline{B}(s)$ is illustrated in Figure 3.3a, along with the largest Nash equilibrium $\bar{s}$ of this game. Similarly, $s \geq \underline{B}(s)$ if and only if each player's strategy is above her best response to the other player's strategy. These strategies and the smallest Nash equilibrium $\underline{s}$ are illustrated in Figure 3.3b.



(a) Largest Nash equilibrium

(b) Smallest Nash equilibrium

Figure 3.3: Applying the Topkis theorem to the partnership game.

3.3 Comparative Statics

Definition 3.9. A **parameterized supermodular game** $(N, (S_i)_{i\in N}, (u_i)_{i\in N}, T)$ is a family of supermodular games with payoff functions that are parameterized by t in some partially ordered set T, such that for each $i \in N$:

(1) $S_i \subseteq \mathbb{R}^{m_i}$ is a lattice and is compact.
(2) $u_i(s_i, s_{-i}, t)$ is continuous in s_i for fixed s_{-i} and t.
(3) $u_i(s_i, s_{-i}, t)$ is quasisupermodular in s_i and satisfies the single crossing property in $(s_i; s_{-i}, t)$.

Theorem 3.10. Suppose $(N, (S_i)_{i\in N}, (u_i)_{i\in N}, T)$ is a parameterized supermodular game, and let $\bar{s}(t)$ and $\underline{s}(t)$ denote the largest and smallest Nash equilibria for each $t \in T$. Then these equilibria are nondecreasing in t.

Proof. For each fixed $t \in T$, define $\overline{B}_i(s_{-i}, t)$, $\underline{B}_i(s_{-i}, t)$, $\overline{B}(s, t)$, and $\underline{B}(s, t)$ as above. We know from Theorem 3.7 that the largest and smallest Nash equilibria are

$$\bar{s}(t) = \sup\{s \in S : s \leq \overline{B}(s, t)\},$$
$$\underline{s}(t) = \inf\{s \in S : \underline{B}(s, t) \leq s\}.$$

By Theorem 2.20, $\overline{B}_i(s_{-i},t)$ and $\underline{B}_i(s_{-i},t)$ are nondecreasing in both s_{-i} and t.[5] Therefore, $\overline{B}(s,t)$ and $\underline{B}(s,t)$ are nondecreasing in both s and t. Thus

$$\begin{aligned} t < t' &\implies \overline{B}(s,t) \le \overline{B}(s,t') \quad \forall s \in S \\ &\implies \{s \in S : s \le \overline{B}(s,t)\} \subseteq \{s \in S : s \le \overline{B}(s,t')\} \\ &\implies \bar{s}(t) \le \bar{s}(t'). \end{aligned}$$

Similarly,

$$\begin{aligned} t < t' &\implies \underline{B}(s,t) \le \underline{B}(s,t') \quad \forall s \in S \\ &\implies \{s \in S : \underline{B}(s,t') \le s\} \subseteq \{s \in S : \underline{B}(s,t) \le s\} \\ &\implies \underline{s}(t) \le \underline{s}(t'). \end{aligned}$$

Thus, the extremal equilibria are nondecreasing in t. □

EXAMPLE 3.11 (PARTNERSHIP GAME, CONTINUED). Suppose the production function in the partnership game depends on a parameter $t > 0$, so $f(s_1, s_2, t) = t s_1^\alpha s_2^\beta$. The utility functions of the two players are therefore now

$$u_1(s_1,s_2,t) = \frac{t s_1^\alpha s_2^\beta}{2} - s_1 \quad \text{and} \quad u_2(s_1,s_2,t) = \frac{t s_1^\alpha s_2^\beta}{2} - s_2.$$

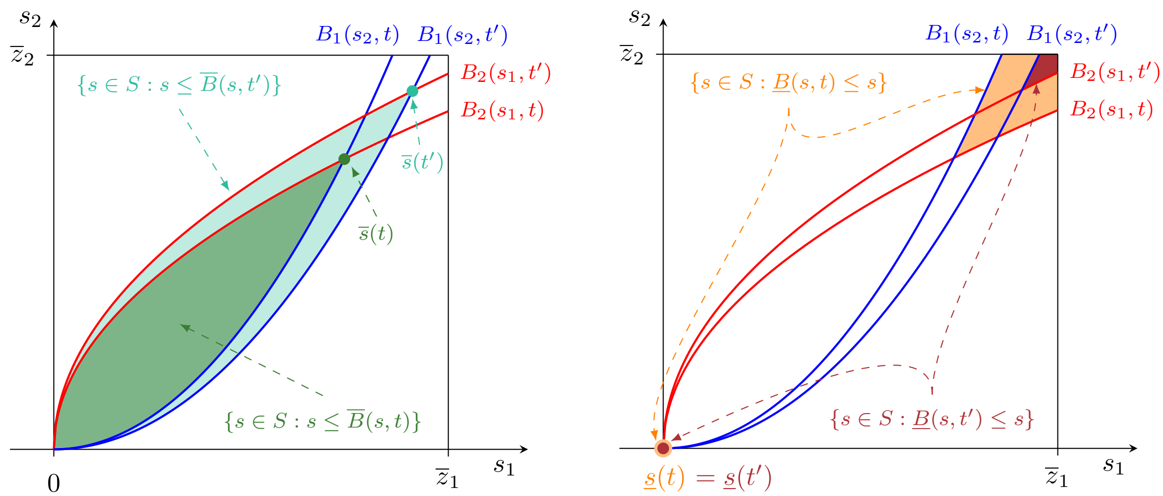
The best responses are now $B_i(s_{-i},t) = \{\underline{B}_i(s_{-i},t)\} = \{\overline{B}_i(s_{-i},t)\}$ where

$$\begin{aligned} \underline{B}_1(s_2,t) = \overline{B}_1(s_2,t) &= \min\left\{ \left(\frac{t\alpha}{2}\right)^{1/(1-\alpha)} s_2^{\beta/(1-\alpha)},\ \bar{z}_1 \right\} \\ \underline{B}_2(s_1,t) = \overline{B}_2(s_1,t) &= \min\left\{ \left(\frac{t\beta}{2}\right)^{1/(1-\beta)} s_1^{\alpha/(1-\beta)},\ \bar{z}_2 \right\}. \end{aligned}$$

Figure 3.4a shows how the set $\{s \in S : s \le \overline{B}(s,t)\}$ becomes larger at t increases, and hence the largest Nash equilibrium $\bar{s}(t)$ also increases. Figure 3.4b illustrates how the set $\{s \in S : \underline{B}(s,t) \le s\}$ becomes smaller as t increases, although in this particular example $\underline{s}(t)$ nonetheless remains constant in t. Note that this is not a contradiction to our results since Theorem 3.10 only implies that the largest and smallest Nash equilibria are nondecreasing, not that they are strictly increasing.

It is important to keep in mind that our comparative statics results for supermodular games do *not* state that the set of Nash equilibria is nondecreasing in t in the strong set order. For instance, in the partnership game example, there are exactly two Nash equilibria, $\underline{s}(t)$ and $\bar{s}(t)$. The first does not change with t, and the second is nondecreasing in t. However,

[5] In this application of the theorem, take the choice variable x for player i to be her strategy s_i and take her "parameter" to be the pair (s_{-i},t).



(a) Largest Nash equilibrium

(b) Smallest Nash equilibrium

Figure 3.4: Nash equilibrium comparative statics in the partnership game ($t < t'$).

the set of Nash equilibria in this example is not nondecreasing in the strong set order, that is, it is not the case that $t' > t$ implies $\{\underline{s}(t'), \bar{s}(t')\} \geq_s \{\underline{s}(t), \bar{s}(t)\}$ (think about why this fails).

The following example illustrates another limitation on the kinds of comparative statics predictions that can be made in supermodular games. It shows that when there are more than two Nash equilibria, the non-extreme equilibria (the Nash equilibria other than $\underline{s}(t)$ and $\bar{s}(t)$) may fail to be nondecreasing.

EXAMPLE 3.12 (SEARCH). Consider an n-player game where the strategy of each player is her search intensity $e_i \in [0,1]$, and let

$$\bar{e}_{-i} = \frac{1}{n-1} \sum_{j \neq i} e_j.$$

Suppose

$$u_i(e_i, e_{-i}, t) = t e_i g(\bar{e}_{-i}) - c(e_i),$$

where $g : [0,1] \to [0,1]$ is increasing and continuously differentiable with $g(0) = 0$, and $c : [0,1] \to \mathbb{R}$ is increasing and continuously differentiable.

$$\frac{\partial^2 u_i}{\partial e_i \partial e_j} = \frac{t g'(\bar{e}_{-i})}{(n-1)} \geq 0 \qquad \text{and} \qquad \frac{\partial^2 u_i}{\partial e_i \partial t} = g(\bar{e}_{-i}) \geq 0,$$

so u_i has increasing differences in $(e_i; e_{-i}, t)$. Thus, this is a parameterized supermodular game. More explicitly, if $c(e_i) = e_i^2/2$ then

$$B_i(e_{-i}, t) = \{t g(\bar{e}_{-i})\}.$$

Since the best responses are nondecreasing and payoffs are symmetric, the largest and smallest equilibria are symmetric (see Exercise 3.1). Symmetric equilibrium satisfy $e_i = tg(e_i)$.

- Graph with three crossings.
- Show middle equilibrium may be decreasing.
- Illustrate best-response dynamics.

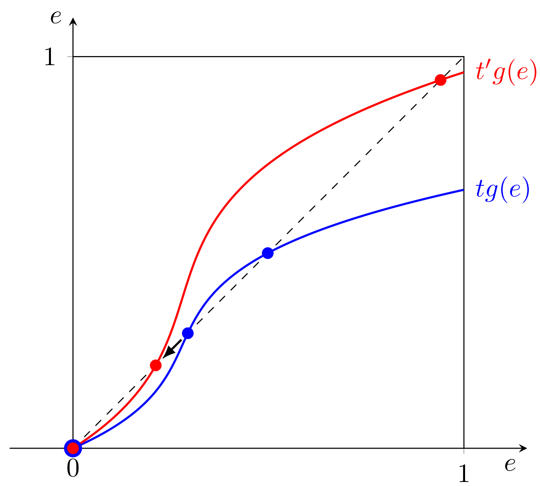


Figure 3.5: Non-monotonicity of non-extreme Nash equilibrium for $t' > t$. Symmetric Nash equilibria satisfy $e = tg(e)$ and $e_i = e$ for all i.

3.4 Iterated Strict Dominance and Rationalizability

- Discuss coordination games as an illustration of why NE is not necessarily a predictive tool absent some sort of learning or adaptive dynamics.
- Although weaker concepts like iterated elimination of strictly dominated strategies and rationalizability might yield weaker predictions, their benefit is that they rely only on common knowledge of rationality.

Definition 3.13. A pure strategy $s_i \in S_i$ for player i is **strictly dominated** by another pure strategy $s'_i \in S_i$ if $u_i(s'_i, s_{-i}) > u_i(s_i, s_{-i})$ for all $s_{-i} \in S_{-i}$. A strategy s_i is **undominated** if it is not dominated by another pure strategy.

Lemma 3.14. Suppose $(N, (S_i)_{i\in N}, (u_i)_{i\in N})$ is a supermodular game. Let $\underline{z}, \bar{z} \in S$ be the smallest and largest strategy profiles, so $\underline{z} \leq s \leq \bar{z}$ for any $s \in S$. If $s_i \ngeq \underline{B}_i(\underline{z}_{-i})$ or $s_i \nleq \overline{B}_i(\bar{z}_{-i})$ then s_i is strictly dominated. Thus, the profiles of undominated strategies for each player are contained in $[\underline{B}(\underline{z}), \overline{B}(\bar{z})] = \{s \in S : \underline{B}(\underline{z}) \leq s \leq \overline{B}(\bar{z})\}$.

The intuition for this result is simple in the case where each player has a single-dimensional strategy space. We begin with a proof for that case before proceeding to the general proof.

Proof—special case of one-dimensional strategy spaces $S_i \subseteq \mathbb{R}$. Since $\underline{B}_i(\underline{z}_{-i})$ is the least best response of player i to the smallest profile $\underline{z}_{-i}$ of other players' strategies, $s_i < \underline{B}_i(\underline{z}_{-i})$ implies that $u_i(\underline{B}_i(\underline{z}_{-i}), \underline{z}_{-i}) > u_i(s_i, \underline{z}_{-i})$. For any strategy profile $s_{-i} \in S_{-i}$ of the other players, since $s_{-i} \geq \underline{z}_{-i}$, single crossing of u_i in $(s_i; s_{-i})$ therefore implies that $u_i(\underline{B}_i(\underline{z}_{-i}), s_{-i}) > u_i(s_i, s_{-i})$. Thus s_i is strictly dominated by $\underline{B}_i(\underline{z}_{-i})$. $\square$

In the case where players have multidimensional strategy spaces, there is a slight complication to this argument since $s_i \not\geq \underline{B}_i(\underline{z}_{-i})$ does not necessarily imply $s_i < \underline{B}_i(\underline{z}_{-i})$ (since s_i could be lower in one dimension but higher in others). Nonetheless, quasisupermodularity of u_i in s_i can be used to prove the result.

Proof—general case. Fix any $i \in N$ and any $s_i \not\geq \underline{B}_i(\underline{z}_{-i})$. Then, it must be the case that s_i is strictly lower than $\underline{B}_i(\underline{z}_{-i})$ in at least one coordinate. Thus, for any $s_{-i} \in S_{-i}$,

$$
\begin{aligned}
s_i \wedge \underline{B}_i(\underline{z}_{-i}) < \underline{B}_i(\underline{z}_{-i}) &\implies u_i(\underline{B}_i(\underline{z}_{-i}), \underline{z}_{-i}) > u_i(s_i \wedge \underline{B}_i(\underline{z}_{-i}), \underline{z}_{-i}) \\
&\implies u_i(s_i \vee \underline{B}_i(\underline{z}_{-i}), \underline{z}_{-i}) > u_i(s_i, \underline{z}_{-i}) \qquad \text{(QSM)} \\
&\implies u_i(s_i \vee \underline{B}_i(\underline{z}_{-i}), s_{-i}) > u_i(s_i, s_{-i}), \qquad \text{(SC)}
\end{aligned}
$$

so s_i is strictly dominated by $s_i \vee \underline{B}_i(\underline{z}_{-i})$. A similar argument applies to any $s_i \not\leq \overline{B}_i(\bar{z}_{-i})$. $\square$

Example 3.15 (Price Competition with Differentiated Products). Suppose firms 1 and 2 simultaneously set prices $p_i \in [0, \bar{z}_i]$. Each firm has a constant marginal cost $c > 0$, and the demand for firm i is

$$D_i(p_i, p_j) = a - bp_i + dp_j,$$

where $a, b, d > 0$ and $a/b > c$.[6] The profit for firm i is therefore

$$\pi_i(p_i, p_j) = (p_i - c)D_i(p_i, p_j),$$

and the best responses are

$$B_i(p_j) = \left\{ \frac{a + bc + dp_j}{2b} \right\}.$$

Figure 3.6 illustrates the best responses of the two firms, along with the points $\underline{B}(\underline{z})$ and $\overline{B}(\bar{z})$. The green rectangle represents the set $[\underline{B}(\underline{z}), \overline{B}(\bar{z})] = \{s \in S : \underline{B}(0,0) \leq s \leq \overline{B}(\bar{z})\}$.

Figure 3.7 illustrates the importance of strategic complementarities in Lemma 3.14, as well as some of the limitations of this result. Figure 3.7a shows that if we remove the

[6] This equation for the demand function permits negative demand when p_i is sufficiently large. However, this is not a serious issue because the inequality $a/b > c$ ensures that the best response of each firm is to set a price where demand is strictly positive.

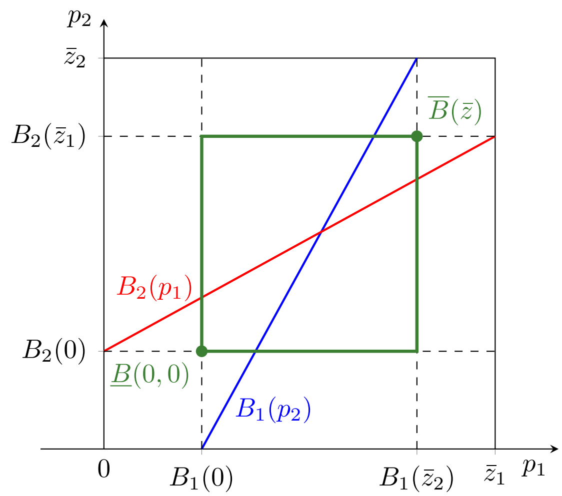
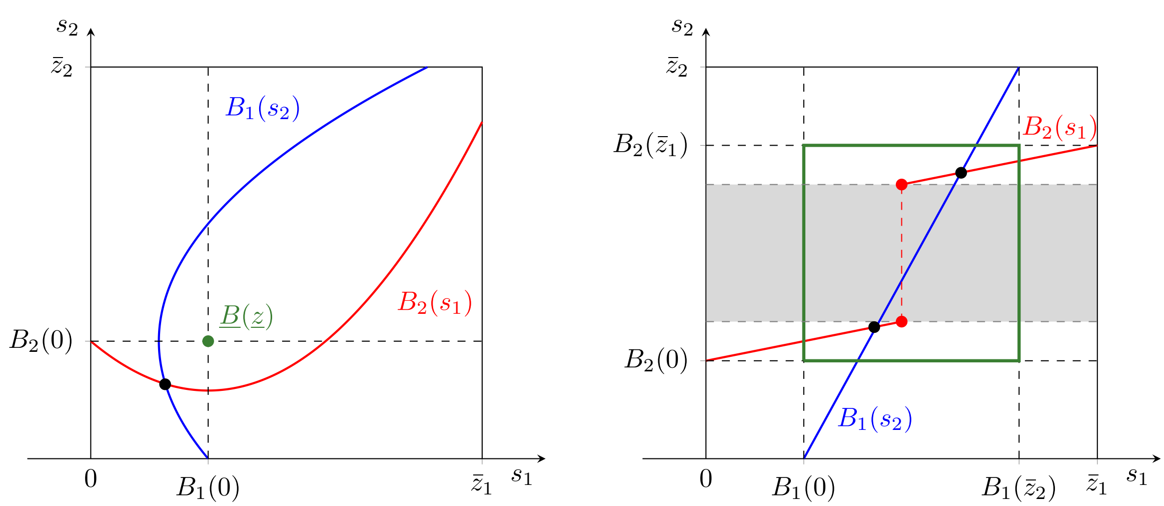


Figure 3.6: Undominated strategies in the price competition game.

assumption of single crossing of u_i in $(s_i; s_{-i})$, and as an extreme allow the best responses to fail to be nondecreasing, then there may be undominated strategies (or even a Nash equilibrium) that lie outside of the set $[\underline{B}(\underline{z}), \overline{B}(\bar{z})]$. Figure 3.7b shows that even when the assumptions of the lemma are satisfied, some dominated strategies may remain in the set $[\underline{B}(\underline{z}), \overline{B}(\bar{z})]$.[7] In other words, while eliminating strategies outside of this set only involves removing strictly dominated strategies, it does not necessarily imply that we have removed *all* strictly dominated strategies.



(a) The lemma works because of complementarity. Without SC, the conclusion may fail.

(b) The lemma does not say that we have eliminated all strictly dominated strategies.

Figure 3.7: More on Lemma 3.14.

[7]Some care is required here. There is a difference between never being a best response to a pure strategy of the other player and being a dominated strategy. However, suppose for the sake of argument that some or all of the strategies in the gray region in Figure 3.7b are indeed strictly dominated for player 2.

The process of **iterated elimination of strictly dominated strategies** proceeds as follows: First, remove all strictly dominated pure strategies for each player i. Denote the remaining (undominated) strategies for player i by S_i^1, and let $S^1 \equiv \prod_{i \in N} S_i^1$ denote the profiles of these undominated strategies. Next, remove all strictly dominated pure strategies in this reduced game where the space of strategy profiles is S^1 to obtain S^2, and so on. A strategy s_i for player i is **serially undominated** if it is not eliminated at any stage of this process of iterated elimination of strictly dominated strategies, that is, $s_i \in S_i^k$ for all $k = 1, 2, \ldots$.

Note that Lemma 3.14 can be applied iteratively to show that the set of strategies surviving two rounds of elimination of strictly dominated strategies are bounded by $\underline{B}^2(\underline{z}) = \underline{B}(\underline{B}(\underline{z}))$ and $\overline{B}^2(\bar{z}) = \overline{B}(\overline{B}(\bar{z}))$, and so on for successive rounds of elimination. That is, $S^k \subseteq [\underline{B}^k(\underline{z}), \overline{B}^k(\bar{z})]$ for all k. What is the limit of this process of iterated elimination? The sequences of largest and smallest best responses $\overline{B}^k(\bar{z})$ and $\underline{B}^k(\underline{z})$, respectively, will converge to precisely the largest and smallest Nash equilibria, provided the game is continuous in the following sense.

Definition 3.16. A supermodular game $(N, (S_i)_{i \in N}, (u_i)_{i \in N})$ is a **continuous supermodular game** if each u_i is continuous in (s_i, s_{-i}).

We have assumed throughout that each u_i is continuous in the player's own strategy s_i. Definition 3.16 imposes the stronger assumption that u_i is jointly continuous in both s_i and the strategies of the other players s_{-i}.

Theorem 3.17 (Milgrom and Roberts (1990)). Suppose $(N, (S_i)_{i \in N}, (u_i)_{i \in N})$ is a continuous supermodular game. Then the set of serially undominated strategy profiles (those that survive iterated elimination of strictly dominated strategies) has largest and smallest elements $\bar{s}$ and $\underline{s}$. Moreover, both of these strategy profiles are Nash equilibria.

Note that the original results in Milgrom and Roberts (1990) assume each u_i is supermodular in s_i and has increasing differences in $(s_i; s_{-i})$. The extension for their result to the ordinal concepts of quasisupermodularity and single crossing presented here is due to Milgrom and Shannon (1994).

Proof. Let $\underline{z}, \bar{z} \in S$ be the smallest and largest strategy profiles, so $\underline{z} \le s \le \bar{z}$ for any $s \in S$. Let $\underline{z}^1 = \underline{B}(\underline{z})$ and $\bar{z}^1 = \overline{B}(\bar{z})$, and inductively define $\underline{z}^{k+1} = \underline{B}(\underline{z}^k)$ and $\bar{z}^{k+1} = \overline{B}(\bar{z}^k)$. By repeated application of Lemma 3.14, any strategy profile that survives k rounds of elimination of strictly dominated strategies must be contained in $[\underline{z}^k, \bar{z}^k] = \{s \in S : \underline{z}^k \le s \le \bar{z}^k\}$.

Observe that $\underline{z}^k$ is an increasing sequence and $\bar{z}^k$ is a decreasing sequence. These claims can be proved by induction. First, note that $\underline{z}^0 \equiv \underline{z} \le \underline{z}^1$ since $\underline{z}$ is the lower bound for all strategies. Second, note that $\underline{z}^{k-1} \le \underline{z}^k$ implies $\underline{z}^k = \underline{B}(\underline{z}^{k-1}) \le \underline{B}(\underline{z}^k) = \underline{z}^{k+1}$ since $\underline{B}$ is nondecreasing. By induction, conclude that $\underline{z}^k \leq \underline{z}^{k+1}$ for all k. A similar argument shows that $\bar{z}^k \geq \bar{z}^{k+1}$ for all k. Since these sequences are bounded and monotone, their limits exist by the monotone convergence theorem. In particular,

$$\underline{s} = \lim_k \underline{z}^k = \sup\{\underline{z}^k : k \in \mathbb{N}\} \qquad \text{and} \qquad \bar{s} = \lim_k \bar{z}^k = \inf\{\bar{z}^k : k \in \mathbb{N}\}.$$

By construction, any profile s of serially undominated strategies must satisfy $\underline{z}^k \leq s \leq \bar{z}^k$ for all $k \in \mathbb{N}$, and hence $\underline{s} \leq s \leq \bar{s}$.

The last step is to show that $\underline{s}$ and $\bar{s}$ are Nash equilibria. Consider first $\underline{s}$. Since $\underline{z}^{k+1} = \underline{B}(\underline{z}^k) \in B(\underline{z}^k)$ for every $k \in \mathbb{N}$, we must have

$$u_i(\underline{z}_i^{k+1}, \underline{z}_{-i}^k) \geq u_i(s_i, \underline{z}_{-i}^k), \quad \forall s_i \in S_i.$$

Taking the limit as $k \to \infty$ and using the continuity of u_i, we obtain

$$u_i(\underline{s}_i, \underline{s}_{-i}) \geq u_i(s_i, \underline{s}_{-i}), \quad \forall s_i \in S_i.$$

Thus $\underline{s}$ is a Nash equilibrium. An analogous argument shows that $\bar{s}$ is a Nash equilibrium. $\square$

Corollary 3.18. A continuous supermodular game with a unique Nash equilibrium is dominance solvable.

We will illustrate these results using the example of price competition with differentiated products introduced above. It may be useful for you to draw similar illustrations for the partnership game and the search game examples.

EXAMPLE 3.19 (PRICE COMPETITION WITH DIFFERENTIATED PRODUCTS, CONTINUED). Consider again the price competition game described in the previous example. As in the proof of Theorem 3.17, define $\underline{z}^1 = \underline{B}(\underline{z})$ and $\bar{z}^1 = \overline{B}(\bar{z})$, and inductively define $\underline{z}^{k+1} = \underline{B}(\underline{z}^k)$ and $\bar{z}^{k+1} = \overline{B}(\bar{z}^k)$. Figure 3.8 illustrates these strategy profiles and the sets $[\underline{z}^k, \bar{z}^k]$ that contain all strategy profiles that survive k rounds of elimination of strictly dominated strategies. In this particular example, there is a unique Nash equilibrium. Notice that the sequences $(\underline{z}^k)_{k \in \mathbb{N}}$ and $(\bar{z}^k)_{k \in \mathbb{N}}$ both converge to this Nash equilibrium. In other words, this game is dominance solvable.

Closely related results showing that best-response dynamics converge to the largest and smallest Nash equilibrium appear in Vives (1990). The preceding results from Milgrom and Roberts (1990), while in some ways quite similar, add the explicit connection to elimination of strictly dominated strategies and rationalizability. Milgrom and Roberts (1990, Theorem 8) also extend the convergence result for best-response dynamics to a broad class of adaptive dynamics, which means that behavior in a wide variety of learning models applied

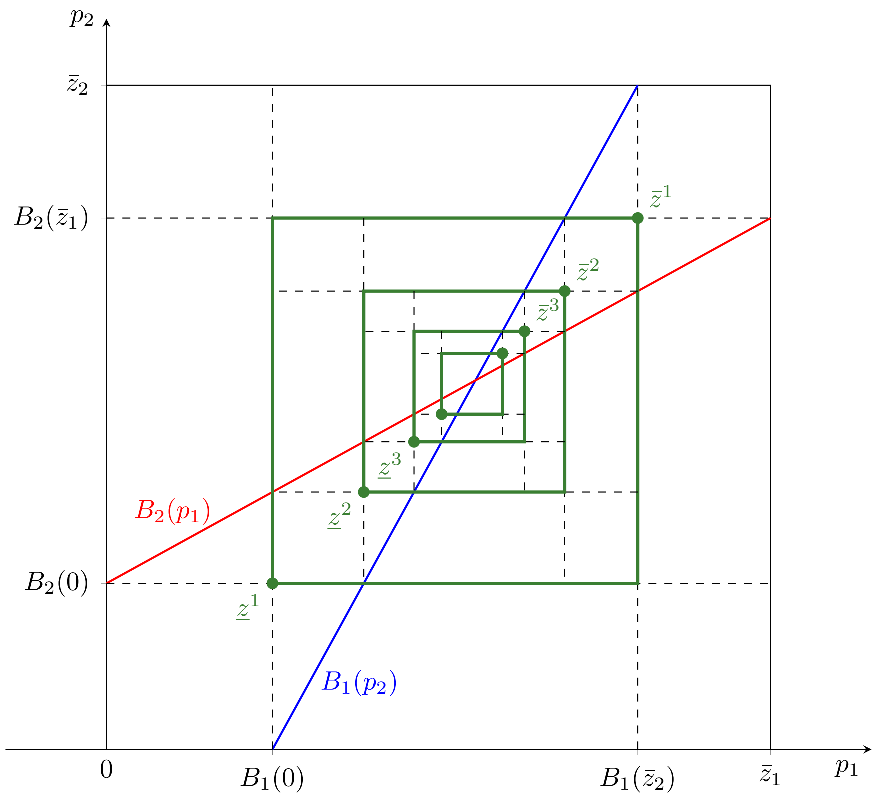


Figure 3.8: Serially undominated strategies in the price competition game.

to a supermodular game should eventually (in the limit) satisfy the bounds described in Theorem 3.17.

3.5 Exercises

3.1 Suppose $(N, (S_i)_{i\in N}, (u_i)_{i\in N})$ is a supermodular game. Suppose the players all have the same strategy space and it is one dimensional, i.e., $S^* \equiv S_1 = \cdots = S_n \subseteq \mathbb{R}$. Suppose the payoffs of the individuals are symmetric. Formally, the game is symmetric if for all permutations $\pi : N \to N$ and for all $s \in S = S^* \times \cdots \times S^*$,

$$u_{\pi(i)}(s_1, \ldots, s_i, \ldots, s_n) = u_i(s_{\pi(1)}, \ldots, s_{\pi(i)}, \ldots, s_{\pi(n)}).$$

This formal definition may be a bit difficult to digest, so consider a simple example to illustrate: Suppose $n = 3$ and $(s_1, s_2, s_3) = (a, b, c)$, and let π be the permutation that swaps 2 and 3 (that is, $\pi(1) = 1, \pi(2) = 3, \pi(3) = 2$). Then, applying this definition to $i = 1$ gives $u_1(a, b, c) = u_1(a, c, b)$. Applying it to $i = 2$ gives $u_3(a, b, c) = u_2(a, c, b)$. Applying it to $i = 3$ gives $u_2(a, b, c) = u_3(a, c, b)$.

(a) Prove that the largest and smallest pure-strategy Nash equilibrium of this game must be symmetric, i.e., each player is playing the same strategy.

(b) Does this imply that every Nash equilibrium of the game is symmetric? Prove or provide a counterexample.

(c) If, in addition, $B_i(s_{-i})$ is singleton-valued (each player i has a unique best response to every pure strategy of the other players), can we make the stronger conclusion that every Nash equilibrium must be symmetric? Prove or provide a counterexample.

(d) What can be said about symmetric games that are not supermodular? Specifically, is there necessarily a largest and smallest Nash equilibrium? Provide a proof or counterexample. If a symmetric game happens to have a largest Nash equilibrium, is that equilibrium necessarily symmetric?

3.2 Consider a Cournot duopoly where each firm chooses a quantity $q_i \in S_i \equiv [0, \bar{q}]$. Suppose each firm has a continuous cost function $C_i : S_i \to \mathbb{R}_+$ and the inverse demand function P is decreasing and twice differentiable with $P'(Q) + \bar{q}P''(Q) < 0$ for all $Q = q_1 + q_2 \in [0, 2\bar{q}]$.

(a) Show that there exist Nash equilibria $(\bar{x}_1, \underline{x}_2)$ and $(\underline{x}_1, \bar{x}_2)$ such that $\underline{x}_i \leq q_i \leq \bar{x}_i$ for any other Nash equilibrium (q_1, q_2).

(b) Suppose that firm 1 receives a government subsidy, receiving $s > 0$ for each unit it sells. Can you say anything definitive about how $\underline{x}_i$ and $\bar{x}_i$ change? Prove any definitive change that you claim or provide a counterexample to illustrate the indeterminacy.

(c) Suppose that we add a constant $\Delta > 0$ to the inverse demand, so that the new price is $\hat{P}(q_1+q_2) = P(q_1+q_2)+\Delta$ for each q_1, q_2. Can you say anything definitive about how $\underline{x}_i$ and $\bar{x}_i$ change? Just discuss whether or not the theorems from the chapter can be applied (a proof or counterexample is not required for this part).

3.3 Consider a two-person partnership game. Simultaneously, each player i invests $s_i \in [0, 1]$, and the payoff of player i is

$$u_i(s_i, s_j, t) = tf(s_1)f(s_2) - C(s_i),$$

where $t \geq 0$ is a parameter, and f and C are continuous and strictly increasing functions with $f(0) > 0$.

(a) Show that the game is supermodular.

(b) Show that the smallest and largest strategies that survive iterated elimination of strictly dominated strategies, as well as the smallest and largest Nash equilibrium strategies, are nondecreasing functions of t.

(c) Give an example showing that the set of Nash equilibria is not nondecreasing in the strong set order.

3.4 Suppose $(N, (S_i)_{i \in N}, (u_i)_{i \in N})$ is a continuous supermodular game. Let $\underline{x} = (\underline{x}_1, \ldots, \underline{x}_n)$

be the smallest strategy profile that survives iterated elimination of strictly dominated strategies. Consider the $n-1$ player game that is created by fixing player n's strategy at $\underline{x}_n$ (so she is no longer a part of the game). Let $\underline{s} = (\underline{s}_1, \ldots, \underline{s}_{n-1})$ be the smallest Nash equilibrium of this new game. Is it the case that $\underline{s}_i = \underline{x}_i$ for all $i \in \{1, \ldots, n-1\}$? Prove or provide a counterexample.

3.5 Consider a firm that has a constant marginal cost of production $c > 0$. The firm chooses a price $p \geq 0$, and the demand for its product is given by a continuously differentiable demand function $D(p)$. Assume that $D(p) > 0$ and $D'(p) \leq 0$ for all $p \geq 0$. The firm treats the marginal cost c as a fixed parameter and chooses a price p to maximize the following profit function:

$$\pi(p, c) = (p - c)D(p).$$

Answer the following:

(a) Based only on the assumptions provided thus far, can we conclude that the set of profit-maximizing prices is nondecreasing in c? If yes, prove it (and also explain the precise sense in which the set of optimal prices is nondecreasing). If not, provide a counterexample.

(b) Suppose for this part of the problem that there is a unique solution for the optimal price, and it is given by the first-order condition $\frac{\partial}{\partial p}\pi(p, c) = 0$. (This is true, for example, if the profit function is strictly concave in the price p.) Show that the optimal price satisfies the equation

$$\frac{p - c}{p} = \frac{1}{\mathcal{E}(p)},$$

where

$$\mathcal{E}(p) = -\frac{p}{D(p)} \frac{\partial D(p)}{\partial p}$$

is the price elasticity of demand. This mark-up relative to price is called the **Lerner index**.

(c) Using the results from part (b), it can be shown that greater elasticity of demand leads to a lower optimal price for the firm. Instead of trying to formalize an argument along these lines based on first-order conditions, let's instead try to apply monotone comparative statics results to this problem. Suppose now that the demand function depends on some parameter $t \in T \subseteq \mathbb{R}$, so demand is given by $D(p, t)$. Suppose also that the price elasticity of demand,

$$\mathcal{E}(p, t) = -\frac{p}{D(p, t)} \frac{\partial D(p, t)}{\partial p},$$

is nondecreasing in the parameter t. Do *not* impose the (overly strong) assumption used in part (b) that the solution is unique. What can we conclude about how a change in t affects the optimal price p for the firm? If there is a tight relationship, prove it. If instead the implications of a change in t are ambiguous, give a careful explanation of why (a counterexample is not required for this part).

Now, consider multiple firms $N = \{1, \dots, n\}$ that compete in prices. The firms simultaneously choose prices $p_i \geq 0$ (you can also assume there is some large upper bound on the set of feasible prices in order to make the strategy spaces compact). The profit function of firm i is

$$\pi_i(p_i, p_{-i}, c_i) = (p_i - c_i) D_i(p_i, p_{-i}),$$

where $c_i > 0$ for each $i \in N$. Assume the demand function D_i takes the following Logit form:

$$D_i(p_i, p_{-i}) = M \frac{e^{-\alpha p_i}}{\sum_{j \in N} e^{-\alpha p_j}}.$$

In this equation, $M > 0$ and $\alpha > 0$ are fixed parameters. Answer the following:

(d) Does a pure-strategy Nash equilibrium exist in this game, and if so, can you say anything definitive about the structure of the set of Nash equilibria of this game? Be precise, and prove any claims that you make.

(e) Suppose the marginal cost c_1 for firm 1 increases. What is the impact on the price p_1 of firm 1 and the prices p_j of firms $j \neq 1$ that can occur in equilibrium? Be precise, and prove any claims that you make.

3.6 Consider a partnership game with two players who invest in a public good project at each date $t \in \{0, 1, 2, \dots, T\}$ without observing each other's previous investments.[8] This game can therefore be treated as simultaneous-move game where the strategy of player i is any function $x_i : \{0, 1, \dots, T\} \to [0, 1]$, where $x_i(t)$ is the investment level of player i at time t. The payoff of player i is

$$u_i(x_1, x_2) = \sum_{t=0}^{T} \delta^t [A f(x_1(t), x_2(t)) - c_i(x_i(t), t)],$$

where $\delta \in (0, 1)$, $A \in (0, 1)$ is a productivity parameter, $f : [0, 1]^2 \to \mathbb{R}$ is a supermodular, increasing, and continuous production function, and c_i is a continuous time-dependent cost function for player i.

(a) Does this game have pure-strategy Nash equilibria $\underline{x}$ and $\bar{x}$ such that for any equilibrium x of the game,

$$\underline{x}_i(t) \leq x_i(t) \leq \bar{x}_i(t)$$

for all $i \in \{1, 2\}$ and $t \in \{0, 1, \dots, T\}$? If yes, prove it. If not, briefly explain why not and give the weakest condition that you can find on f and c_i that is sufficient for this to be true. Prove that your condition is sufficient.

[8] Note that $T \in \mathbb{N}$ is a fixed terminal time period, not a parameter as in some of our theorems.

(b) Maintain any assumptions that were needed for the existence of a largest and smallest Nash equilibrium in part (a). Treat $A \in (0,1)$ as a parameter of the game. Show that if $A' > A$ then the largest and smallest equilibrium are nondecreasing:

$$\underline{x}_i(t, A') \geq \underline{x}_i(t, A) \quad \text{and} \quad \bar{x}_i(t, A') \geq \bar{x}_i(t, A), \quad \forall i, t.$$

3.7 (Based on Bulow, Geanakoplos, and Klemperer (1985)) Consider two firms, 1 and 2, and two markets, A and B. Firm 1 is a monopolist in market B and a duopolist with firm 2 in market A. The firms simultaneously choose their quantities, so a strategy for firm 1 is a pair (q_1^A, q_1^B) and a strategy for firm 2 is a quantity q_2^A. Suppose the inverse demand in market A is

$$P^A(q_1^A, q_2^A) = \alpha^A - \beta^A q_1^A - \beta^A q_2^A$$

for $\alpha^A, \beta^A > 0$, and the inverse demand in market B is

$$P^B(q_1^B) = \alpha^B - \beta^B q_1^B$$

for $\alpha^B, \beta^B > 0$. The firms have nonnegative and twice continuously differentiable cost functions $C_1(q_1^A, q_1^B)$ and $C_2(q_2^A)$.

(a) To gain intuition, consider a special case of the general model where

$$\begin{aligned} P^A(q_1^A, q_2^A) &= 100 - q_1^A - q_2^A \\ P^B(q_1^B) &= 50 \\ C_1(q_1^A, q_1^B) &= \frac{1}{2}(q_1^A + q_1^B)^2 \\ C_2(q_2^A) &= \frac{1}{2}(q_2^A)^2. \end{aligned}$$

Solve for the Cournot-Nash equilibrium quantities is this example.

(b) Keep all of the same assumptions as in part (a), except suppose there is a positive shock to demand in market B that changes the inverse demand function to $P^B(q_1^B) = 60$. Find the new Cournot-Nash equilibrium. In particular, what happens to firm 1's equilibrium quantity in market A?

(c) Move back to the general linear demand specification with parameters $\alpha^A, \beta^A, \alpha^B, \beta^B$ and cost functions $C_1(q_1^A, q_1^B)$ and $C_2(q_2^A)$. Suppose there is a positive shock to demand in market B in the form of an increase in α^B.

- Under what conditions (if any) will this lead to a (weakly) larger equilibrium quantity for firm 1 in market B?
- Under what conditions (if any) will this lead to a (weakly) larger equilibrium quantity for firm 1 in market A?
- Under what conditions (if any) will this lead to a (weakly) smaller equilibrium quantity for firm 1 in market A?

(You can assume the Cournot-Nash equilibrium is unique to simplify the statement of your answer.)

3.8 Consider a multi-market oligopoly problem with two firms, 1 and 2, and three markets, A, B, and C. Firm 1 sells in markets A and B, firm 2 sells in markets B and C, and no other firms sell in these markets. Thus firm 1 is a monopoly in market A, the firms are duopolists in market B, and firm 2 is a monopoly in market C. The firms simultaneously choose quantities, with firm 1 choosing $q_1^A, q_1^B \in [0, \bar{q}]$ and firm 2 choosing $q_2^B, q_2^C \in [0, \bar{q}]$. Make the following assumptions about the inverse demand functions:

- Inverse demand in market A is $P^A(q_1^A, t) = t + f(q_1^A)$, where $t \geq 0$ is a parameter and f is continuously differentiable and strictly decreasing function.
- Inverse demand in market B is $P^B(q_1^B, q_2^B) = \alpha - (q_1^B + q_2^B)$.
- Inverse demand in market C is a continuously differentiable and strictly decreasing function $P^C(q_2^C)$.

You can also assume that each of these functions is strictly positive even when both firms produce the maximum quantities $\bar{q}$ in each of the markets they serve. Assume that the cost functions $C_1(q_1^A, q_1^B)$ and $C_2(q_2^B, q_2^C)$ are twice continuously differentiable. Answer each of the following based solely on the information provided above together whether the additional information provided about the cost function(s) in each part. If you invoke unnecessary assumptions to draw conclusions about any variables, you can still receive partial but not full credit. To simplify the problem, you can also assume throughout this problem that the Cournot-Nash equilibrium is unique.[9]

(a) Suppose C_1 is additively separable, meaning $C_1(q_1^A, q_1^B) = C_1^A(q_1^A) + C_1^B(q_1^B)$ for twice continuously differentiable functions C_1^A and C_1^B. Based on this information, for the equilibrium values of each of q_1^A, q_1^B, q_2^B, q_2^C, specify whether the quantity is weakly increasing in t, weakly decreasing in t, constant in t, or whether there is not enough information to tell if any of these three conditions hold. Provide precise arguments in support of any definitive conclusions you draw, and provide a brief explanation in support of any cases where you say there is insufficient information.[10]

(b) Suppose $C_1(q_1^A, q_1^B)$ is supermodular in (q_1^A, q_1^B). Based on this information, for the equilibrium values of each of q_1^A, q_1^B, q_2^B, q_2^C, specify whether the quantity is weakly increasing in t, weakly decreasing in t, constant in t, or whether there is not enough information to tell if any of these three conditions hold. Provide

[9] Just to be sure there is no confusion: The information provided does *not* allow us to conclude that there is a unique equilibrium. I'm saying that you can (without penalty) make that additional assumption in order to simplify your answers.

[10] In other words, if you say that we cannot determine in which direction a particular equilibrium quantity will move, just explain why our theorems do not apply. You do *not* need to provide counterexamples to show that the quantity could move in either direction.

precise arguments in support of any definitive conclusions you draw, and provide a brief explanation in support of any cases where you say there is insufficient information.

(c) Suppose $C_1(q_1^A, q_1^B) = (q_1^A)^2 + (q_1^B)^2 + (q_1^A + q_1^B)^{1/2}$ and $C_2(q_2^B, q_2^C) = (q_2^B)^2 + (q_2^C)^2 + (q_2^B + q_2^C)^2$. Based on this information, for the equilibrium values of each of q_1^A, q_1^B, q_2^B, q_2^C, specify whether the quantity is weakly increasing in t, weakly decreasing in t, constant in t, or whether there is not enough information to tell if any of these three conditions hold. Provide precise arguments in support of any definitive conclusions you draw, and provide a brief explanation in support of any cases where you say there is insufficient information.

3.9 This problem considers an arms race between two countries. In the static version of the game, each player $i = 1, 2$ chooses a level of arms $x_i \in [0, \bar{x}]$ and receives a utility of

$$u_i(x_i, x_{-i}) = B(x_i - x_{-i}) - C(x_i).$$

Assume that $B : [-\bar{x}, \bar{x}] \to \mathbb{R}$ is a twice continuously differentiable, (weakly) increasing, and (weakly) concave function. Assume $C : \mathbb{R} \to \mathbb{R}$ is a twice continuously differentiable function of any shape. Intuitively, the function C captures the cost of investing in a level of arms x_i, and $B(x_i - x_{-i})$ captures the payoffs associated with the resulting conflict (which only depends on the difference in the level of arms of the two countries).[11] Answer the following:

(a) Under the assumptions given, is this a supermodular game? If yes, then list the conditions required by the definition of a supermodular game and confirm that they are satisfied, and state what this implies about the existence and structure of the set of Nash equilibria of this game. If no, then explain which condition(s) in the definition of a supermodular game could be violated and what additional assumptions would ensure that this is a supermodular game.

(b) Consider now some micro-foundations for the function B described above. For each of the following definitions of B, you will be asked to illustrate the function and explain whether the function is twice continuously differentiable, (weakly) increasing, and (weakly) concave.

i. Suppose

$$B(z) = \begin{cases} 1 & \text{if } z > 0 \\ 0 & \text{if } z = 0 \\ -1 & \text{if } z < 0, \end{cases}$$

where z is the difference in arms levels, that is, $z = x_i - x_{-i}$. Draw a graph of this function $B : [-\bar{x}, \bar{x}] \to \mathbb{R}$. Which of the conditions listed above does this definition of B satisfy and which does it violate?

[11] Note that this type of game is traditionally referred to as an "arms race" following the literal interpretation suggested above, but the game applies equally well to R&D races, political lobbying, and a host of other applications.

ii. Suppose $B(z) = \Pr[z > \varepsilon] - \Pr[z < \varepsilon]$, where ε is a uniformly distributed random variable on the interval $[-\alpha, \alpha]$ and $z = x_i - x_{-i}$ is again the difference in arms levels. (Note that when $\alpha = 0$, the variance of ε is zero and we have the same definition of B as in the previous part of the problem.) Draw a graph of this function $B : [-\bar{x}, \bar{x}] \to \mathbb{R}$. For what range of values of α, if any, does this definition of B satisfy all of the conditions listed above? Explain.

Now, consider a two-period version of this game. Let x_i be the level of arms of country $i = 1, 2$ at time $t = 1$, and let y_i be the level of arms of country i at time $t = 2$. These levels of arms are determined by the level of investment I_i^t of country i at each time t according to the following equations:

$$\begin{aligned} x_i &= I_i^1 \\ y_i &= (1-\delta)x_i + I_i^2, \end{aligned}$$

where $\delta \in (0, 1)$ is the rate of depreciation or the arms stockpile. Intuitively, fraction $(1 - \delta)$ of the investment in time $t = 1$ remains in time $t = 2$ and the investment in time $t = 2$ is added to this amount. The utility of each player $i = 1, 2$ is now given by

$$u_i = B(x_i - x_{-i}) - C(I_i^1) + B(y_i - y_{-i}) - C(I_i^2).$$

Note that we can write $I_i^2 = y_i - (1-\delta)x_i$ and therefore write the utility of each player solely as a function of the armament levels x_1, y_1, x_2, y_2:

$$u_i(x_i, y_i, x_{-i}, y_{-i}) = B(x_i - x_{-i}) - C(x_i) + B(y_i - y_{-i}) - C(y_i - (1-\delta)x_i). \quad (3.1)$$

Suppose, as before, that the armament levels in each period must satisfy the constraint $x_i, y_i \in [0, \bar{x}]$. That is, the upper bound $\bar{x}$ applies to the level of arms that can be stored in each period (not to the amount I_i^t that can be invested). Answer the following:

(c) Suppose that we write this as a game of choosing arms levels, as in Equation (3.1), so the strategy of each player i is (x_i, y_i).[12] Assume that $B : [-\bar{x}, \bar{x}] \to \mathbb{R}$ is a twice continuously differentiable, (weakly) increasing, and (weakly) concave function. Assume that $C : \mathbb{R} \to \mathbb{R}$ is twice continuously differentiable, (weakly) increasing, and (weakly) convex. Under these assumptions, is this a supermodular game? Explain.

(d) Maintaining the same assumptions on B and C that were given in part (c), does this game have a unique Nash equilibrium? If yes, argue carefully why this is the

[12] Note that by writing the strategies in this way, we are implicitly assuming that y_i cannot be conditioned on x_j; that is, the period 2 investment strategy of country i is a single numerical value and not a function that can depend on the period 1 investment by country j. This assumption might be sensible if either (i) the investments levels must be chosen well in advance and cannot be adjusted based on observed investments by the other country, or (ii) the investment level of the other country cannot be observed until the end of period 2. Under either of these assumptions, this interaction can be modeled as a static game, as we have done here. If, instead, the period 2 investment of county i could depend on the period 1 investment by country j, how would the game change? What solution concept would you use?

case. If no, explain why not and provide the weakest sufficient conditions on B and/or C that you can find that ensure that there is a unique Nash equilibrium. Prove your conditions are sufficient. (Note: You are not required to find necessary conditions, but finding weaker sufficient conditions and avoiding unnecessary assumptions will earn more points.)

(e) Are there sufficient conditions on B and/or C such that this game dominance solvable? (Reminder: A game is dominance solvable if the processes of iterated elimination of strictly dominance strategies leaves only a *single* strategy profile.) If yes, provide the weakest sufficient conditions that you can find such that this is true. (Note: As in the previous part, you are not required to find necessary conditions, but finding weaker sufficient conditions and avoiding unnecessary assumptions will earn more points.) If no, explain why this is not possible.

Chapter 4

Comparing Risk and Risk Aversion

Contents

References and Assigned Readings

Primary readings:

(1) Mas-Colell, Whinston, and Green (1995), Section 6.D (FOSD and SOSD) (recommended).
(2) Mas-Colell, Whinston, and Green (1995), Section 6.C (risk aversion and comparative risk aversion) (recommended).
(3) Pratt (1964) (recommended).

Additional references:

(1) Machina and Pratt (1997) (updated treatment of Rothschild and Stiglitz (1970, 1972) characterizations of SOSD).
(2) Gollier (2001)

4.1 First-Order Stochastic Dominance

4.1.1 Definition and Equivalent Characterizations

Definition 4.1. A **cumulative distribution function** is a function $F : \mathbb{R} \to \mathbb{R}$ that is nondecreasing, right-continuous, and satisfies

$$\lim_{x \to -\infty} F(x) = 0 \quad \text{and} \quad \lim_{x \to \infty} F(x) = 1.$$

As the following definition formalizes, one cumulative distribution function first-order stochastically dominates another if every individual with a nondecreasing Bernoulli utility function prefers the first to the second.

Definition 4.2. Suppose F and G are cumulative distribution functions on $\mathbb{R}$. We say F **first-order stochastically dominates** G, denoted $F \geq_{FOSD} G$, if

$$\int u(x)\, dF(x) \geq \int u(x)\, dG(x). \tag{4.1}$$

for every nondecreasing function $u : \mathbb{R} \to \mathbb{R}$ for which both integrals are defined.[1]

[1]Throughout these lecture notes, integration is in the sense of Lebesgue-Stieltjes. If you are not already familiar with the basics of measure theory and these different integration concepts, you might find the following YouTube videos useful for a brief overview:

- Riemann vs Lebesgue integral: `https://youtu.be/PGPZ0P1PJfw`
- Lebesgue-Stieltjes measures: `https://youtu.be/IsmgLGVpLpQ`

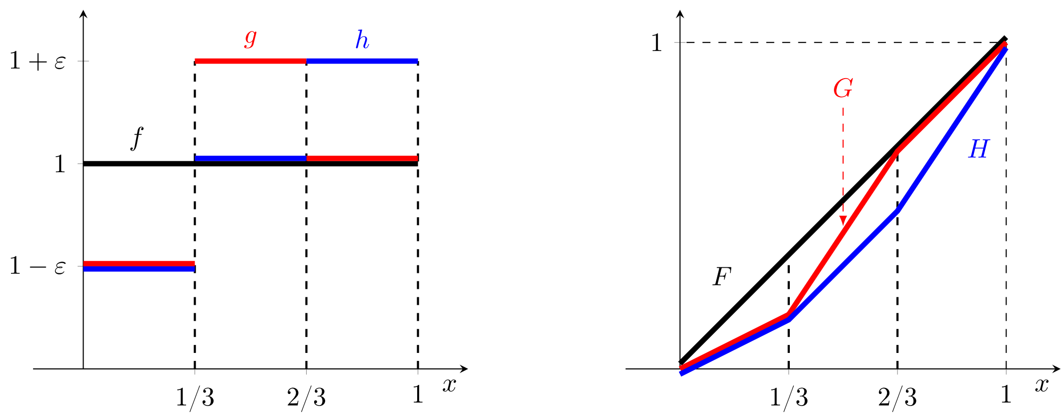
Theorem 4.3. The distribution F first-order stochastically dominates G if and only if $F(x) \leq G(x)$ for all $x \in \mathbb{R}$.

The following examples illustrate this stochastic order.

Example 4.4. Suppose F, G, H have support $[0,1]$ and admit the following density functions:

$$f(x) = 1 \qquad g(x) = \begin{cases} 1-\varepsilon & \text{if } 0 \leq x \leq \frac{1}{3} \\ 1+\varepsilon & \text{if } \frac{1}{3} < x \leq \frac{2}{3} \\ 1 & \text{if } \frac{2}{3} < x \leq 1 \end{cases} \qquad h(x) = \begin{cases} 1-\varepsilon & \text{if } 0 \leq x \leq \frac{1}{3} \\ 1 & \text{if } \frac{1}{3} < x \leq \frac{2}{3} \\ 1+\varepsilon & \text{if } \frac{2}{3} < x \leq 1 \end{cases}$$

Figure 4.1 illustrates these density functions and the corresponding cumulative distribution functions F, G, and H, respectively. Notice that $H \geq_{FOSD} G \geq_{FOSD} F$.



(a) Probability density functions

(b) Cumulative distribution functions

Figure 4.1: Illustration of Example 4.4.

Example 4.5 (Uniform distributions). Suppose F is the uniform distribution on the interval $[0, a]$. That is,

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ x/a & \text{if } 0 \leq x \leq a \\ 1 & \text{if } a < x. \end{cases}$$

Similarly, suppose G is the uniform distribution on the interval $[0, b]$. It is each to see that if $a \geq b$, then $F(x) \leq G(x)$ for all $x \in \mathbb{R}$ and hence $F \geq_{FOSD} G$.

Example 4.6 (Binary discrete distributions). Suppose the distribution F assigns probability $\alpha \in [0,1]$ to the point $a > 0$ and assigns probability $1-\alpha$ to the point 0. Then,

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1-\alpha & \text{if } 0 \leq x < a \\ 1 & \text{if } a \leq x. \end{cases}$$

Similarly, suppose G assigns probability $\beta \in [0,1]$ to the point $b > 0$ and assigns probability $1-\beta$ to the point 0. If $\alpha \geq \beta$ and $a \geq b$, then $F \geq_{FOSD} G$. Notice that this illustrates at least two ways of obtaining a first-order stochastic dominance increase of the distribution G: First, one could fix the probabilities by taking $\alpha = \beta$ and increase the upper mass point from b to some larger $a > b$. Second, one could fix the supports by taking $a = b$ and increase the probability of this outcome from β to some larger $\alpha > \beta$. Of course, making both changes simultaneously also leads to a first-order stochastic dominance increase.

There are different ways of representing uncertainty about outcomes. Sometimes working with a cumulative distribution function is most convenient. Other times, working with random variables is easier. It is important to understand the connections between these different ways of representing uncertainty and the implications of various stochastic orders in each of these domains. Suppose X is a random variable defined on a probability space $(\Omega, \mathcal{F}, P)$. That is, $X : \Omega \to \mathbb{R}$. The **distribution** of X is defined by

$$F_X(x) = P(\{\omega \in \Omega : X(\omega) \leq x\}).$$

For two random variables X and Y, we write $X \geq Y$ if $X(\omega) \geq Y(\omega)$ for all $\omega \in \Omega$.

Corollary 4.7. Suppose X and Y are random variables defined on a probability space $(\Omega, \mathcal{F}, P)$. If $X \geq Y$, then $F_X \geq_{FOSD} F_Y$.

Proof. If $X(\omega) \geq Y(\omega)$ for all $\omega \in \Omega$, then for any $x \in \mathbb{R}$,

$$\{\omega \in \Omega : X(\omega) \leq x\} \subseteq \{\omega \in \Omega : Y(\omega) \leq x\}.$$

Therefore,

$$F_X(x) = P(\{\omega \in \Omega : X(\omega) \leq x\}) \leq P(\{\omega \in \Omega : Y(\omega) \leq x\}) = F_Y(x).$$

Since this is true for any $x \in \mathbb{R}$, Theorem 4.3 implies $F_X \geq_{FOSD} F_Y$. □

The converse of this corollary is not true. It is possible to have random variables X and Y where $X(\omega) < Y(\omega)$ for some ω, yet $F_X \geq_{FOSD} F_Y$. The following example illustrates.

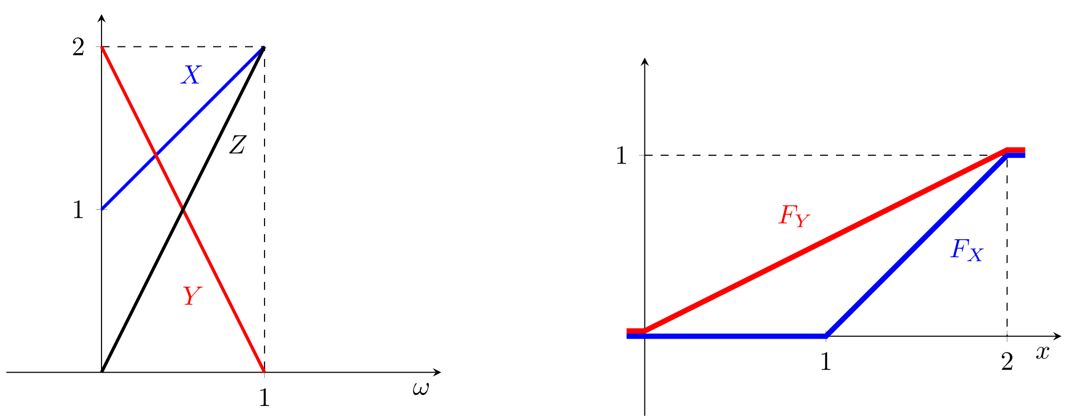
Example 4.8. Suppose $\Omega = [0,1]$ and P is the uniform distribution. Let

$$X(\omega) = 1 + \omega \qquad \text{and} \qquad Y(\omega) = 2 - 2\omega.$$

Then, $Y(\omega) > X(\omega)$ for all $\omega < 1/3$, but

$$F_X(x) = \begin{cases} 0 & \text{if } x < 1 \\ x - 1 & \text{if } 1 \le x < 2 \\ 1 & \text{if } 2 \le x \end{cases} \qquad F_Y(x) = \begin{cases} 0 & \text{if } x < 0 \\ \frac{x}{2} & \text{if } 0 \le x < 2 \\ 1 & \text{if } 2 \le x \end{cases}$$

Thus, $F_X(x) \le F_Y(x)$ for all x, so $F_X \ge_{FOSD} F_Y$ even though $X \not\ge Y$. However, note that we can construct another random variable that has the same distribution as Y and that is pointwise dominated by X. For example, define $Z(\omega) = 2\omega$. Then, Y and Z have the same distribution, $F_Y = F_Z$, and therefore $F_X \ge_{FOSD} F_Z$. In addition, $X \ge Z$.



(a) Random variables (b) Cumulative distribution functions

Figure 4.2: Illustration of Example 4.8.

As hinted at by the previous example, a partial converse of Corollary 4.7 is true. Specifically, if $G \ge_{FOSD} H$ then there exists a probability space $(\Omega, \mathcal{F}, P)$ and random variables X and Y on this space such that $F_X = G$, $F_Y = H$, and $X \ge Y$. In fact, constructing such random variables in one approach to proving Theorem 4.3, as we show next.

4.1.2 Proof of Characterization Theorem (optional)

For the rest of this section, let $(\Omega, \mathcal{F}, P)$ be the probability space consisting of the uniform measure on the unit interval: Let $\Omega = (0,1)$, let $\mathcal{F}$ be the Borel σ-algebra on $(0,1)$, and let P be the uniform (Lebesgue) measure.

Lemma 4.9. *If $F(x) \le G(x)$ for all $x \in \mathbb{R}$, then there exist random variables $X : (0,1) \to \mathbb{R}$ and $Y : (0,1) \to \mathbb{R}$ such that the distribution of X is F, the distribution of Y is G, and $X \ge Y$.*

Proof. Define random variables $X : (0,1) \to \mathbb{R}$ and $Y : (0,1) \to \mathbb{R}$ by

$$X(\omega) = \inf\{x \in \mathbb{R} : F(x) \geq \omega\},$$
$$Y(\omega) = \inf\{x \in \mathbb{R} : G(x) \geq \omega\}.$$

We first show that the distribution of X is F and the distribution of Y is G. To see this, note that

$$\begin{aligned} X(\omega) \leq x &\iff x \geq \inf\{x' \in \mathbb{R} : F(x') \geq \omega\} \\ &\iff x \in \{x' \in \mathbb{R} : F(x') \geq \omega\} \qquad \text{(since } F \text{ is right continuous)} \\ &\iff \omega \leq F(x). \end{aligned}$$

Thus for any $x \in \mathbb{R}$,

$$F_X(x) = P(\{\omega \in (0,1) : X(\omega) \leq x\}) = P(\{\omega \in (0,1) : \omega \leq F(x)\}) = F(x).$$

A similar argument shows that $F_Y(x) = G(x)$.

Next, to see that $X \geq Y$, recall that F and G were assumed to satisfy $F(x) \leq G(x)$ for all $x \in \mathbb{R}$. Therefore, for any $\omega \in (0,1)$,

$$\{x \in \mathbb{R} : F(x) \geq \omega\} \subseteq \{x \in \mathbb{R} : G(x) \geq \omega\},$$

and hence $X(\omega) \geq Y(\omega)$. Since this is true for all ω, we have $X \geq Y$. □

Using this lemma, we can now easily prove Theorem 4.3.

Proof of Theorem 4.3. We first prove that if $F(x) \leq G(x)$ for all $x \in \mathbb{R}$, then Equation (4.1) is satisfied for every nondecreasing function $u : \mathbb{R} \to \mathbb{R}$. To show this, define random variables X and Y as in Lemma 4.9, so $F_X = F$, $F_Y = G$, and $X \geq Y$. For any nondecreasing function u, we must therefore have $u(X(\omega)) \geq u(Y(\omega))$ for all $\omega \in (0,1)$. Thus,

$$\int_\Omega u(X(\omega))\, dP(\omega) \geq \int_\Omega u(Y(\omega))\, dP(\omega).$$

Since F is the distribution of X and G is the distribution of Y, this is equivalent to

$$\int_{-\infty}^{\infty} u(x)\, dF(x) \geq \int_{-\infty}^{\infty} u(x)\, dG(x).$$

Hence, Equation (4.1) is satisfied.

To prove the converse, suppose it is not the case that $F(x) \leq G(x)$ for all $x \in \mathbb{R}$. This implies there is some $\bar{x} \in \mathbb{R}$ such that $F(\bar{x}) > G(\bar{x})$. Consider the nondecreasing function

$u : \mathbb{R} \to \mathbb{R}$ defined by

$$u(x) = \begin{cases} 0 & \text{if } x \leq \bar{x} \\ 1 & \text{if } x > \bar{x}. \end{cases}$$

Then,

$$\begin{aligned} \int u(x)\, dF(x) &= (0)F(\bar{x}) + (1)(1 - F(\bar{x})) \\ &< (0)G(\bar{x}) + (1)(1 - G(\bar{x})) = \int u(x)\, dG(x), \end{aligned}$$

and hence Equation (4.1) is violated. □

4.2 Monotone Likelihood Ratio Order

To study the monotone likelihood ratio order, we will consider two classes of distributions:

(1) F is continuously distributed with a density function f, so that

$$F(x) = \int_{-\infty}^{x} f(t)\, dt.$$

(2) F is discretely distributed with either finite support $\{x_1, \ldots, x_n\}$ or a countably infinite support $\{x_1, x_2, \ldots\}$ and a discrete density function f, so that

$$F(x) = \sum_{x_i \leq x} f(x_i).$$

Note that this restriction is not without loss of generality, as there are many distributions that do not admit density functions. Nonetheless, many distributions of interest will fall into one of these two cases.

Definition 4.10. Suppose F and G are cumulative distribution functions on $\mathbb{R}$, and suppose these distributions have (either discrete or continuous) density functions f and g, respectively. Then F dominates G in the **monotone likelihood ratio order**, denoted $F \geq_{MLR} G$ (or $f \geq_{MLR} g$), if

$$f(x')g(x) \geq f(x)g(x'), \quad \forall x' > x.$$

When g is strictly positive, then the condition in this definition can be written as

$$\frac{f(x')}{g(x')} \geq \frac{f(x)}{g(x)}, \quad \forall x' > x,$$

that is, $f(x)/g(x)$ is nondecreasing in x.

EXAMPLE 4.11. Suppose F, G, H are defined as in Example 4.4. That is,

$$f(x) = 1 \qquad g(x) = \begin{cases} 1-\varepsilon & \text{if } 0 \le x \le \frac{1}{3} \\ 1+\varepsilon & \text{if } \frac{1}{3} < x \le \frac{2}{3} \\ 1 & \text{if } \frac{2}{3} < x \le 1 \end{cases} \qquad h(x) = \begin{cases} 1-\varepsilon & \text{if } 0 \le x \le \frac{1}{3} \\ 1 & \text{if } \frac{1}{3} < x \le \frac{2}{3} \\ 1+\varepsilon & \text{if } \frac{2}{3} < x \le 1 \end{cases}$$

It is easy to see that $F \not\geq_{MLR} G$ and $G \not\geq_{MLR} H$. However, $F \geq_{MLR} H$.

EXAMPLE 4.12 (UNIFORM DISTRIBUTIONS, CONTINUED). Suppose as in Example 4.5 that F is the uniform distribution on the interval $[0, a]$ and G is the uniform distribution on the interval $[0, b]$. It is not difficult to show that if $a \geq b$, then $F \geq_{MLR} G$.

EXAMPLE 4.13 (BINARY DISCRETE DISTRIBUTIONS, CONTINUED). Suppose as in Example 4.6 that the distribution F assigns probability $\alpha \in [0, 1]$ to the point $a > 0$ and assigns probability $1 - \alpha$ to the point 0. Similarly, G assigns probability $\beta \in [0, 1]$ to the point $b > 0$ and assigns probability $1 - \beta$ to the point 0. We make two observations about these distributions:

- It is easy to see that if $a = b$ and $\alpha \geq \beta$, then $F \geq_{MLR} G$.
- However, if $a \neq b$ and $\alpha, \beta \in (0, 1)$, then $F \not\geq_{MLR} G$, even if $a > b$ and $\alpha > \beta$. To see this, note that $b > 0$, yet

$$f(b)g(0) = 0(1-\beta) \not\geq (1-\alpha)\beta = f(0)g(b),$$

violating the definition of the MLR order.

Contrasting Examples 4.4 and 4.6 with their counterparts in Examples 4.11 and 4.13, we see that first-order stochastic dominance does not imply dominance in the monotone likelihood ratio order. That is, FOSD is not strong enough to imply MLR dominance. However, the opposite is true: The MLR order is stronger (more restrictive) than the FOSD order, as the following lemma demonstrates.

Lemma 4.14. If $F \geq_{MLR} G$ then $F \geq_{FOSD} G$.

Proof. We will prove the result for the case where F and G are continuously distributed. The proof for discrete distributions is analogous. Fix any x. Then for any $t' \geq x \geq t$, $F \geq_{MLR} G$ implies that $f(t')g(t) \geq f(t)g(t')$. Integrating over $t \in (-\infty, x]$, we have

$$f(t')G(x) = \int_{-\infty}^{x} f(t')g(t)\,dt \geq \int_{-\infty}^{x} f(t)g(t')\,dt = F(x)g(t'), \quad \forall t' \geq x.$$

Then, integrating over $t' \in [x, \infty)$, we have

$$(1 - F(x))G(x) = \int_x^\infty f(t')G(x)\,dt' \geq \int_x^\infty F(x)g(t')\,dt' = (1 - G(x))F(x).$$

This is only possible if $F(x) \leq G(x)$. Since x was arbitrary, this completes the proof. $\square$

Lemma 4.14 shows that the monotone likelihood ratio order is more restrictive than the first-order stochastic dominance order. However, it may not be clear exactly what additional structure is imposed by the MLR order. The following theorem demonstrates precisely what the MLR order requires beyond first-order stochastic dominance.

Theorem 4.15. Suppose F and G admit continuous and strictly positive density functions f and g, respectively.[2] Then, the following statements are equivalent:

(1) $F \geq_{MLR} G$.
(2) The conditional distribution of F first-order stochastically dominates the conditional distribution of G on every interval $[a, b]$. Formally, for every $a < b$, we have

$$\frac{F(x) - F(a)}{F(b) - F(a)} \leq \frac{G(x) - G(a)}{G(b) - G(a)} \qquad \forall x \in [a, b].$$

Proof. We first prove that 2 implies 1. Fix any $a < x < b$. Note that

$$\frac{F(x) - F(a)}{F(b) - F(a)} = \frac{F(x) - F(a)}{F(b) - F(x) + F(x) - F(a)} = \frac{1}{\frac{F(b)-F(x)}{F(x)-F(a)} + 1},$$

and similarly for G. Therefore, condition 2 implies

$$\frac{F(b) - F(x)}{F(x) - F(a)} \geq \frac{G(b) - G(x)}{G(x) - G(a)},$$

and hence

$$\frac{F(b) - F(x)}{G(b) - G(x)} \geq \frac{F(x) - F(a)}{G(x) - G(a)}.$$

This is true for any $a < x < b$. Therefore, applying the same inequality for $x < b < x'$, we have

$$\frac{F(x') - F(b)}{G(x') - G(b)} \geq \frac{F(b) - F(x)}{G(b) - G(x)}.$$

[2] Assuming strictly positive density functions ensures that the probability assigned by both F and G to any interval $[a, b]$ with $b > a$ is strictly positive. The theorem and its proof can be extended to relax this assumption, but with slightly more cumbersome arguments.

Combining these two inequalities, we have

$$\frac{F(x')-F(b)}{G(x')-G(b)} \geq \frac{F(b)-F(x)}{G(b)-G(x)} \geq \frac{F(x)-F(a)}{G(x)-G(a)},$$

and hence

$$\frac{f(x')}{g(x')} = \lim_{b \to x'} \frac{\frac{F(x')-F(b)}{x'-b}}{\frac{G(x')-G(b)}{x'-b}} \geq \lim_{a \to x} \frac{\frac{F(x)-F(a)}{x-a}}{\frac{G(x)-G(a)}{x-a}} = \frac{f(x)}{g(x)}.$$

Since this is true for any $x' > x$, we have $F \geq_{MLR} G$.

To prove that 1 implies 2, fix any $a < b$. Let F^* and G^* denote the conditional distributions on this interval, and let f^* and g^* denote the conditional densities. That is,

$$F^*(x) = \begin{cases} 0 & \text{if } x < a \\ \frac{F(x)-F(a)}{F(b)-F(a)} & \text{if } a \leq x \leq b \\ 1 & \text{if } x > b \end{cases} \qquad \text{and} \qquad f^*(x) = \begin{cases} 0 & \text{if } x < a \\ \frac{f(x)}{F(b)-F(a)} & \text{if } a \leq x \leq b \\ 0 & \text{if } x > b, \end{cases}$$

with G^* and g^* defined similarly. Note that $F \geq_{MLR} G$ implies that $F^* \geq_{MLR} G^*$ (this is easily verified using the formulas for f^* and g^* and the definition of the MLR order). Therefore, by Lemma 4.14, $F^* \geq_{FOSD} G^*$, which establishes condition 2. □

4.3 Second-Order Stochastic Dominance

4.3.1 Mean-Preserving Reductions in Risk

Theorem 4.16. For any cumulative distribution functions F and G on $[a, b]$, the following are equivalent:

(1) There exist a pair of random variables X and ε on some probability space $(\Omega, \mathcal{F}, P)$ such that F is the cumulative distribution function of X, G is the cumulative distribution function of $X + \varepsilon$, and $\mathbb{E}[\varepsilon | X] = 0$.[3]

(2) For all $t \in [a, b]$,

$$\int_a^t F(x)\, dx \leq \int_a^t G(x)\, dx \qquad \text{and} \qquad \int_a^b F(x)\, dx = \int_a^b G(x)\, dx.$$

(3) For every concave function $u : [a, b] \to \mathbb{R}$,

$$\int u(x)\, dF(x) \geq \int u(x)\, dG(x).$$

Note that these conditions imply that F and G have the same mean. This can be seen

[3]Using less precise but perhaps more descriptive notation, this means $\mathbb{E}[\varepsilon | X = x] = 0$ for all x.

in a number of ways. Using the first condition, the law of iterated expectations implies

$$\mathbb{E}[X+\varepsilon] = \mathbb{E}\big[X + \mathbb{E}[\varepsilon|X]\big] = \mathbb{E}[X].$$

This can also be seen from the second condition. Since $F(a) = G(a) = 0$ (since F and G have bounded support, we can always choose a so that this is true) and $F(b) = G(b) = 1$, integration by parts yields

$$\int_a^b F(x)\,dx = \big[xF(x)\big]_{x=a}^b - \int_a^b x\,dF(x) = b - \int_a^b x\,dF(x),$$

and likewise for G. Thus

$$\int_a^b F(x)\,dx = \int_a^b G(x)\,dx \iff \int_a^b x\,dF(x) = \int_a^b x\,dG(x).$$

Finally, that F and G have the same mean also follows from the third condition applied to the functions $u(x) = x$ and $u(x) = -x$.

Definition 4.17. If any of the conditions in Theorem 4.16 hold, we say that F differs from G by a **mean-preserving reduction in risk** and, equivalently, that G differs from F by a **mean-preserving increase in risk**.

Remark 4.18. I will write $F \geq_{MPRR} G$ if F differs from G by a mean-preserving reduction in risk. However, note that this is not established notation. In fact, even the terminology in Definition 4.17 (which comes from Machina and Pratt (1997)) is not fully established. In the operations research literature on stochastic orders, Condition 3 in Theorem 4.16 is referred to as dominance in the concave order.[4] Specifically, F is larger than G in the **concave order**, denoted $F \geq_{cv} G$, if

$$\int u(x)\,dF(x) \geq \int u(x)\,dG(x)$$

for all concave $u : [a,b] \to \mathbb{R}$.[5] Thus writing $F \geq_{MPRR} G$ (as we will in this course) is equivalent to writing $F \geq_{cv} G$ (as in the operations research literature).

Example 4.19. Consider the probability space $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$ with $P(\omega) = 1/4$ for

[4] For example, see Shaked and Shanthikumar (2007, page 109).

[5] The operations research typically focuses on convexity rather than concavity, just as it usually focuses on minimization rather than maximization. Therefore, it is even more common to see the **convex order**, which is defined as above but for all convex $u : [a,b] \to \mathbb{R}$. Note that F is larger than G in the concave order if and only if G is larger than F in the convex order.

each $\omega \in \Omega$. Suppose the random variables X and ε are defined as follows:

$$X(\omega_1) = X(\omega_2) = 1 \quad \text{and} \quad X(\omega_3) = X(\omega_4) = 10$$
$$\varepsilon(\omega_1) = -1, \quad \varepsilon(\omega_2) = 1, \quad \varepsilon(\omega_3) = -2, \quad \varepsilon(\omega_4) = 2.$$

Let $Y = X + \varepsilon$, so

$$Y(\omega_1) = 0, \quad Y(\omega_2) = 2, \quad Y(\omega_3) = 8, \quad Y(\omega_4) = 12.$$

Since the partition generated by the random variable X is $\{\{\omega_1, \omega_2\}, \{\omega_3, \omega_4\}\}$, the conditional expectation of ε given X is

$$\begin{aligned}\mathbb{E}[\varepsilon|X](\omega) &= \begin{cases} \frac{1}{2}\varepsilon(\omega_1) + \frac{1}{2}\varepsilon(\omega_2) & \text{if } \omega = \omega_1, \omega_2 \\ \frac{1}{2}\varepsilon(\omega_3) + \frac{1}{2}\varepsilon(\omega_4) & \text{if } \omega = \omega_3, \omega_4 \end{cases} \\ &= 0, \quad \forall \omega \in \Omega.\end{aligned}$$

Thus, letting F_X and F_Y denote the distributions of the random variables X and Y, respectively, F_X is a mean preserving reduction in risk of F_Y by condition 1 in Theorem 4.16. These cumulative distribution functions are illustrated below in Figure 4.3. This figure also illustrates condition 2 in Theorem 4.16. For the value t indicated in the figure, the difference in the value of the integrals $\int_0^t F_Y(x)\,dx - \int_0^t F_X(x)\,dx$ is the difference between the red shaded areas and the blue shaded areas, which is positive. It is easy to see that this is true for any $t \in [0, 12]$, and moreover, the integrals are equal for $t = 12$.

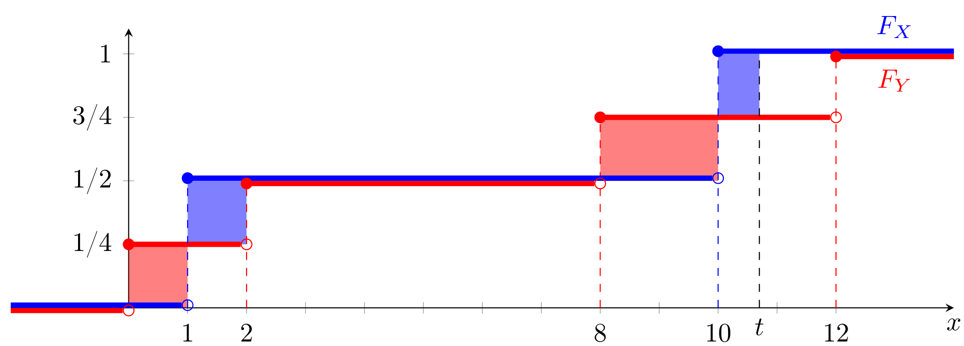


Figure 4.3: Illustration of Example 4.19 and the integral condition from Theorem 4.16.

Rothschild and Stiglitz (1970) pointed out some of the limitations of mean-variance analysis and instead suggested using mean-preserving increases in risk as a comparative measure of variability. Note that this notion of an increase in risk is stronger than an increase in variance. The following lemma formalizes this claim; the proof is left as an exercise.

Lemma 4.20. If $F \geq_{MPRR} G$, then $\text{Var}[X] \leq \text{Var}[Y]$ for any random variables X and Y with distributions F and G, respectively.

While Rothschild and Stiglitz (1970) are credited with bringing the general notions of reductions in risk discussed in this section into the forefront of the economics literature, earlier formal results more general than Theorem 4.16 can be found in the statistics literature, most notably, in Blackwell (1951, 1953). We will discuss Blackwell's seminal papers on the comparison of statistical experiments in a later chapter. Rothschild and Stiglitz (1972) later acknowledged Blackwell's work on the subject after they became aware of it.

4.3.2 Second-Order Stochastic Dominance

While Rothschild and Stiglitz (1970) focused on random variables with the same mean in their analysis of increasing risk, these ideas can be extended to random variables with different means, as the following theorem and definition show.

Theorem 4.21. For any cumulative distribution functions F and G on $[a, b]$, the following are equivalent:

(1) There exist a pair of random variables X and ε on some probability space $(\Omega, \mathcal{F}, P)$ such that F is the cumulative distribution function of X, G is the cumulative distribution function of $X + \varepsilon$, and $\mathbb{E}[\varepsilon|X] \leq 0$.[6]

(2) For all $t \in [a, b]$,
$$\int_a^t F(x)\,dx \leq \int_a^t G(x)\,dx.$$

(3) For every nondecreasing and concave function $u : [a, b] \to \mathbb{R}$,
$$\int u(x)\,dF(x) \geq \int u(x)\,dG(x).$$

Definition 4.22. In any of the conditions in Theorem 4.21 hold, we say that F **second-order stochastically dominates** G, denoted $F \geq_{SOSD} G$.

REMARK 4.23. In the operations research literature on stochastic orders, $F \geq_{SOSD} G$ is sometimes referred to as F being larger than G in the **increasing concave order** and written as $F \geq_{icv} G$.[7] However, referring to this order as second-order stochastic dominance is, by far, more common in economics.

EXAMPLE 4.24. Consider a slight modification of Example 4.19 where X is defined just as

[6] Equivalently, there exist random variables X and Y such that F is the distribution of X, G is the distribution of Y, and $\mathbb{E}[Y|X] \leq X$.

[7] For example, Müller and Stoyan (2002, page 16) or Shaked and Shanthikumar (2007, page 181).

before, but $\varepsilon(\omega_3) = -4$ (rather than -2) and hence $Y(\omega_3) = 6$. That is,

$$X(\omega_1) = X(\omega_2) = 1 \quad \text{and} \quad X(\omega_3) = X(\omega_4) = 10$$
$$\varepsilon(\omega_1) = -1, \quad \varepsilon(\omega_2) = 1, \quad \varepsilon(\omega_3) = -4, \quad \varepsilon(\omega_4) = 2$$
$$Y(\omega_1) = 0, \quad Y(\omega_2) = 2, \quad Y(\omega_3) = 6, \quad Y(\omega_4) = 12.$$

We now have

$$\begin{aligned}\mathbb{E}[\varepsilon|X](\omega) &= \begin{cases} \frac{1}{2}\varepsilon(\omega_1) + \frac{1}{2}\varepsilon(\omega_2) & \text{if } \omega = \omega_1, \omega_2 \\ \frac{1}{2}\varepsilon(\omega_3) + \frac{1}{2}\varepsilon(\omega_4) & \text{if } \omega = \omega_3, \omega_4 \end{cases} \\ &= \begin{cases} 0 & \text{if } \omega = \omega_1, \omega_2 \\ -1 & \text{if } \omega = \omega_3, \omega_4. \end{cases}\end{aligned}$$

Therefore, letting F_X and F_Y denote the distributions of X and Y, respectively, F_X is *not* a mean preserving reduction in risk of F_Y, but $F_X \geq_{SOSD} F_Y$. Therefore, not every individual with a concave Bernoulli utility function u will prefer X to Y, but every individual with a concave and nondecreasing Bernoulli utility function will. Figure 4.4 illustrates the cumulative distribution functions F_X and F_Y. It is easy to see that condition 2 from Theorem 4.21 is satisfied. In particular,

$$\int_0^t F_X(x)\,dx \leq \int_0^t F_Y(x)\,dx, \quad \forall t \in [0, 12],$$

but now with a strict inequality at $t = 12$.

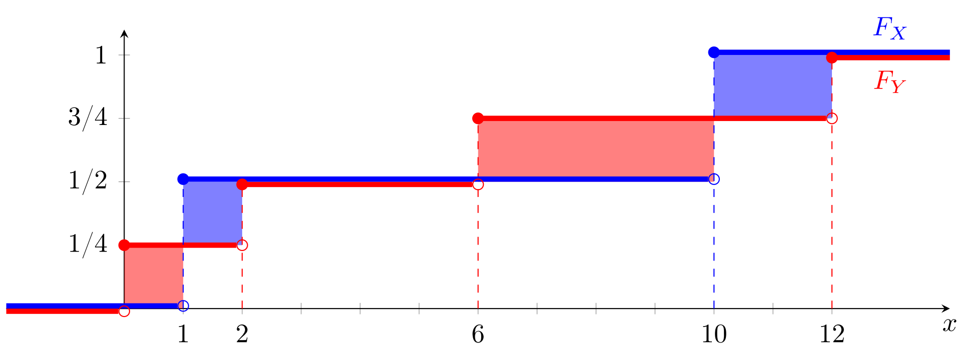


Figure 4.4: Illustration of Example 4.24 and the integral condition from Theorem 4.21.

By condition 3 in Theorem 4.21, it is immediate that the first-order stochastic dominance order is stronger (more restrictive) than the second-order stochastic dominance order. The relationships between the orders described so far are summarized as follows:

$$F \geq_{MLR} G \implies F \geq_{FOSD} G \implies F \geq_{SOSD} G,$$
$$F \geq_{MPRR} G \implies F \geq_{SOSD} G.$$

The fact that the second-order stochastic dominance order permits increases in mean as well as decreases in variability might suggest that it is in some sense a combination of first-order stochastic dominance with a mean-preserving decrease in risk. The following lemma shows that this intuition is accurate.

Lemma 4.25. Fix any distributions F and G on $[a, b]$. Then $F \geq_{SOSD} G$ if and only if there exists a distribution H such that

$$F \geq_{MPRR} H \geq_{FOSD} G.$$

Proof. If $F \geq_{MPRR} H \geq_{FOSD} G$, then $F \geq_{SOSD} H \geq_{SOSD} G$. By the transitivity of the second-order stochastic dominance order, $F \geq_{SOSD} G$.

To prove the converse, suppose $F \geq_{SOSD} G$. By Theorem 4.21, there exist random variables X and ε such that F is the distribution of X, G is the distribution of $Y = X + \varepsilon$, and $\mathbb{E}[\varepsilon|X] \leq 0$. Define a random variable Z by

$$Z = Y - \mathbb{E}[\varepsilon|X] = X + \varepsilon - \mathbb{E}[\varepsilon|X],$$

and let H be the distribution of Z. Then $Z \geq Y$ since $\mathbb{E}[\varepsilon|X] \leq 0$ and therefore $H \geq_{FOSD} G$ by Corollary 4.7. Define a random variable $\hat{\varepsilon} = \varepsilon - \mathbb{E}[\varepsilon|X]$. By construction,

$$\mathbb{E}[\hat{\varepsilon}|X] = \mathbb{E}\big[\varepsilon - \mathbb{E}[\varepsilon|X]\big|X\big] = \mathbb{E}[\varepsilon|X] - \mathbb{E}[\varepsilon|X] = 0,$$

where the second equality follows from the law of iterated expectations. Note that

$$Z = Y - \mathbb{E}[\varepsilon|X] = X + \hat{\varepsilon},$$

and hence $F \geq_{MPRR} H$ by Theorem 4.16. $\square$

4.4 Comparative Measures of Risk Aversion

In this section, we briefly review the definitions and various characterizations of risk aversion. We begin with absolute risk aversion, where "absolute" means we are asking when a single individual should be classified as risk averse. We next explore comparative risk aversion, where "comparative" means we are asking when one individual should be classified as more risk aversion than another.

These definitions can be stated either in terms of lotteries (meaning probability distributions) or random variables. We will focus on random variables in our definitions, since this will make it more convenient to define decreasing (or increasing) absolute and relative risk aversion. Keep in mind that if a preference $\succsim$ is defined on the space of random variables, it induces a preference over lotteries, and vice versa.[8] Throughout this section, we are implic-

[8]Formally, if $\succsim$ is defined on the space of probability distributions, then it induces a preference $\succsim'$ over

itly assuming that the preference $\succsim$ is defined for gambles over some interval of real numbers (i.e., the domain of the Bernoulli utility function u is an interval), and we are considering only random variables X that take values in this interval.

Definition 4.26. An individual with risk preference $\succsim$ is **risk averse** if she prefers to receive the expected value of a random variable (for certain) to holding the random variable, that is, $\mathbb{E}[X] \succsim X$ for any random variable X.

Before describing some equivalent characterizations of risk aversion, we begin by defining a few useful objects. If $\succsim$ is a preference over random variables, the **certainty equivalent** of a random variable X, denoted $CE_{\succsim}(X)$, is the deterministic consumption value such the individual is indifferent between this sure consumption and the original random variable:

$$CE_{\succsim}(X) \sim X.$$

Under mild regularity properties, a risk preference has a certainty equivalent for every gamble. The **risk premium** of X, denoted $RP_{\succsim}(X)$, is the maximum amount that the individual would be willing to pay to avoid the risk in X and instead receive its expected value as a certain outcome:

$$X \sim \big(\mathbb{E}[X] - RP_{\succsim}(X)\big).$$

There is an obvious relationship between the risk premium and the certainty equivalent: $RP_{\succsim}(X) = \mathbb{E}[X] - CE_{\succsim}(X)$.

Theorem 4.27. Consider a risk preference $\succsim$ that has an expected-utility representation with a continuous and increasing Bernoulli utility function u. The following are equivalent:

(1) $\succsim$ is risk averse.
(2) $CE_{\succsim}(X) \leq \mathbb{E}[X]$ for any random variable X.
(3) $RP_{\succsim}(X) \geq 0$ for any random variable X.
(4) $\int u(x)\, dF(x) \leq u(\int x\, dF(x))$ for any distribution F.
(5) u is concave.

The proof of this result is straightforward and is left as an exercise.

With the definition of absolute risk aversion in hand, we now proceed to compare the risk aversion of two individuals.

---

random variables, where we set $X \succsim' Y$ if and only if $F_X \succsim F_Y$. Conversely, if we start with a preference $\succsim$ over random variables such that $X \sim Y$ whenever $F_X = F_Y$ (think about why this property is important), then it induces a preference $\succsim'$ over lotteries, where we set $G \succsim' H$ if and only if there exist random variables X and Y such that $F_X = G$, $F_Y = H$, and $X \succsim Y$.

Definition 4.28. One risk preference $\succsim_1$ is **more risk averse** than another $\succsim_2$ if for any random variable X and any deterministic consumption value x,

$$X \succsim_1 x \implies X \succsim_2 x.$$

Definition 4.29. Given a twice differentiable Bernoulli utility functions u for money, the Arrow-Pratt **coefficient of absolute risk aversion** at x is

$$A(x) = -\frac{u''(x)}{u'(x)}.$$

Theorem 4.30. Consider any pair of risk preferences $\succsim_1$ and $\succsim_2$ that have expected-utility representations with continuous and increasing Bernoulli utility functions u_1 and u_2, respectively. The following are equivalent:

(1) $\succsim_1$ is more risk averse than $\succsim_2$.
(2) $w + X \succsim_1 w \implies w + X \succsim_2 w$ for any deterministic wealth level w and any random variable X.
(3) $CE_{\succsim_1}(X) \leq CE_{\succsim_2}(X)$ for any random variable X.
(4) $RP_{\succsim_1}(X) \geq RP_{\succsim_2}(X)$ for any random variable X.
(5) $\int u_1(x)\, dF(x) \geq u_1(\bar{x}) \implies \int u_2(x)\, dF(x) \geq u_2(\bar{x})$ for any distribution F and riskless outcome $\bar{x}$.
(6) u_1 is a concave transformation of u_2, that is, there exists an increasing and concave function ϕ such that $u_1(x) = \phi(u_2(x))$.
(7) $A_1(x) \geq A_2(x)$ for all x.

We will not prove this result here, although the proof is not difficult (see Pratt (1964) for the original treatment, or Mas-Colell, Whinston, and Green (1995) or Kreps (2013) for a textbook treatment).

4.5 Exercises

4.1 Consider the subset $\{(p_1, p_3) : p_1 \geq 0, p_3 \geq 0, p_1 + p_3 \leq 1\}$ of $\mathbb{R}^2$. (We are using p_3 to denote the second coordinate for reasons that will become clear shortly.) This set is sometimes referred to as the Marschak-Machina triangle and can be used to graphically represent probability distributions on any set of three prizes $X = \{x_1, x_2, x_3\}$. Specifically, assume the prizes are real numbers with $x_3 > x_2 > x_1$ and let p_1 be the probability of x_1, $p_2 = 1 - p_1 - p_3$ be the probability of x_2, and p_3 be the probability of x_3.[9] Draw this triangle, and choose an arbitrary point $p = (p_1, p_3)$ somewhere in the

[9] Graphically, the bottom-right corner on the horizontal axis is the distribution that gives x_1 with probability 1, the corner at the origin is the distribution that gives x_2 with probability 1, and the upper corner

interior of the triangle.

(a) Illustrate in your graph all the distributions in your triangle that first-order stochastically dominate (FOSD) this distribution p. Provide equations or calculations to support your answer.
(b) On a new graph or on your previous graph, illustrate all the distributions in your triangle that monotone likelihood ratio (MLR) dominate this distribution p. Provide equations or calculations to support your answer.
(c) To avoid clutter, start a new graph for this part. Draw the Marschak-Machina triangle and your arbitrary distribution p again. Illustrate all the distributions in your triangle that are a mean-preserving reduction in risk (MPRR) of this distribution p. Provide equations or calculations to support your answer.
(d) On a new graph or on your previous graph, illustrate all the distributions in your triangle that second-order stochastically dominate (SOSD) this distribution p. Provide equations or calculations to support your answer.
(e) What relationship between these stochastic orders is indicated by your figures? That is, indicate which of these four stochastic orders implies which of the others (in the sense that q dominates p with respect to the first order implies q dominates p with respect to the second).

4.2 In what follows, suppose X, Y, and Z are bounded random variables, and let F_X, F_Y, and F_Z denote the cumulative distribution functions of these random variables. Also, let F_{X+Z} and F_{Y+Z} denote the cumulative distributions of the random variables $X+Z$ and $Y+Z$, respectively. If you find it convenient, you are free to assume that these random variables admit density functions f_X, f_Y, f_Z, f_{X+Z}, and f_{Y+Z} (you are not required to make this assumption, and it is not needed to solve the problem, but you are free to make it without penalty). Answer the following:

(a) Suppose X, Y, and Z are independent random variables, and suppose that $F_X \geq_{FOSD} F_Y$, that is, F_X first-order stochastically dominates F_Y. Can we conclude that $F_{X+Z} \geq_{FOSD} F_{Y+Z}$? If yes, prove it. If not, provide a counterexample.
(b) You are now asked to determine if your answer to part (a) depends on the assumption that the random variables are independent. That is, do **not** assume that X, Y, and Z are independent. If $F_X \geq_{FOSD} F_Y$, can we conclude that $F_{X+Z} \geq_{FOSD} F_{Y+Z}$? If yes, prove it. If not, provide a counterexample. (If you answered "no" to part (a), you do not need to construct a new counterexample for this part—you can simply refer to your previous counterexample.)
(c) Suppose X, Y, and Z are independent random variables, and suppose that $F_X \geq_{MPRR} F_Y$, that is, F_X is a mean-preserving reduction in risk of F_Y. Can we conclude that $F_{X+Z} \geq_{MPRR} F_{Y+Z}$? If yes, prove it. If not, provide a counterexample.
(d) Is this part of the problem, do **not** assume that X, Y, and Z are independent.

on the vertical axis is the distribution that gives x_3 with probability 1.

Suppose that $F_X \neq F_Y$, that is, X and Y do not have the same distribution. Is it possible to have $F_X \geq_{MPRR} F_Y$ and also have $F_{X+Z} \leq_{MPRR} F_{Y+Z}$? In other words, loosely speaking, is it possible to have X be less risky than Y and yet have $X+Z$ be more risky than $Y+Z$? If yes, provide an example where this occurs. If not, prove that this is impossible.

4.3 Suppose an individual has wealth w that she is allocating between a riskless asset with gross return R^f and a risky asset with gross return $\tilde{R}$. The individual chooses an investment $\alpha \in \mathbb{R}$ in the risky asset, which yields a stochastic consumption of

$$(w-\alpha)R^f + \alpha\tilde{R} = wR^f + \alpha(\tilde{R} - R^f).$$

Note that we are permitting $\alpha < 0$, which corresponds to a short sale of the asset, and $\alpha > w$, which corresponds to leveraging her investment in the risky asset by borrowing at the risk free rate. The individual has a Bernoulli utility function u that twice differentiable, strictly increasing, and concave.

(a) Show that if $\mathbb{E}[\tilde{R}] > R^f$, then every solution (if there is a solution) must be strictly positive ($\alpha > 0$). [Hint: This proof can be based on first-order conditions and does not require any sort of comparative statics analysis.]
(b) Show that if $\mathbb{E}[\tilde{R}] = R^f$, then $\alpha = 0$ is a solution. (Could there be others? What conditions would ensure a unique solution?)
(c) Show that if $\mathbb{E}[\tilde{R}] < R^f$, then every solution (if there is a solution) must involve a short sale ($\alpha < 0$).
(d) Suppose now that the individual is constrained from making short sales and must have $\alpha \geq 0$. Show that if $\mathbb{E}[\tilde{R}] < R^f$, then $\alpha = 0$ is the unique solution.

4.4 Suppose an individual has wealth w and faces a loss of amount L with probability $\pi \in (0,1)$. Suppose she has the option to purchase insurance coverage. She can choose any level of coverage $\alpha \in [0, L]$ at a price $p > 0$ per unit of coverage. That is, she pays a premium of $p\alpha$ for her coverage and receives a payment of α in the event of a loss. Model all of these events as transpiring in a single period; that is, there is a single period in which the insurance premium is paid, the loss is realized (in the event that it occurs), and the insurance company makes a payout (if the loss is realized). Suppose the individual has a Bernoulli utility function u that is twice differentiable, strictly increasing, and strictly concave.

(a) We say that insurance is **actuarially fair** if the price of the policy is equal to the expected payout by the insurance company ($p = \pi$). Show that if insurance is actuarially fair, the individual will purchase full insurance ($\alpha = L$).
(b) We say that insurance is **actuarially unfair** if the price of the policy is strictly greater than the expected payout by the insurance company ($p > \pi$). Show that if insurance is actuarially unfair, the individual will purchase less than full insurance

$(\alpha < L)$.

(c) Suppose we relax the assumption that u is differentiable. Can you construct an example of a Bernoulli utility function u that is strictly increasing and concave, but not differentiable everywhere, such that for some value of w, L, π, p the individual would be willing to purchase full coverage even if the insurance is actuarially unfair? Briefly (and without any formal details) explain why your example is not "generic," in the sense that it relies very specific parameter values. (Although we assumed strict concavity of u in this problem, you are permitted to use an example that is concave but not strictly concave for this part.)

4.5 This problem explores some of the mechanics of comparative risk aversion and decreasing (constant, increasing) absolute and relative risk aversion. The following (equivalent) definition of comparative risk aversion will be convenient for this problem: Individual 1 is more risk averse than individual 2 if for any random variable X and any deterministic wealth level w,

$$w + X \succsim_1 w \implies w + X \succsim_2 w.$$

Answer the following: (I am not expecting you to prove anything from first principles. The goal is for you to see how to apply the theorems and definitions from this chapter in a clever way that allows you to address the issues in this problem.)

(a) Think now about a single individual, and suppose this individual is an expected-utility maximizer with an increasing and twice differentiable Bernoulli utility function u. We say that the individual exhibits **decreasing absolute risk aversion (DARA)** if $w + X \succsim w$ implies $w' + X \succsim w'$ for every $w' > w$ and every random variable X. Prove that the individual exhibits DARA if and only if $A(w)$ is nonincreasing in w, where $A(w) = -\frac{u''(w)}{u'(w)}$ is the coefficient of absolute risk aversion.

(b) We similarly say that the individual exhibits **increasing absolute risk aversion (IARA)** if $w + X \succsim w \implies w' + X \succsim w'$ for every $w' < w$ and every random variable X, and that the individual exhibits **constant absolute risk aversion (CARA)** if $w + X \succsim w \iff w' + X \succsim w'$ for all w, w'. Prove that the individual exhibits IARA (CARA) if and only if $A(w)$ is nondecreasing (constant) in w.

(c) We say that an individual exhibits **decreasing relative risk aversion (DRRA)** if $w + X \succsim w$ implies $kw + kX \succsim kw$ for every $k > 1$ and every wealth level w and random variable X. We can also define the Arrow-Pratt **coefficient of relative risk aversion** at w as

$$R(w) = -w\frac{u''(w)}{u'(w)}$$

Prove that the individual exhibits DRRA if and only if $R(w)$ is nonincreasing in w. (We can similarly define increasing and constant relative risk aversion in the obvious ways and show that they are equivalent to $R(w)$ being nondecreasing and constant, respectively.)

(d) Using the equivalences established in the previous parts of the problem, can you determine the relationship between DARA and DRRA (i.e., does one imply the other)?

Chapter 5

Monotone Comparative Statics Under Risk

Contents

References and Assigned Readings

Primary readings:

(1) Athey (2002) (optional).

Additional references:

(1) Gollier (2001)
(2) Rothschild and Stiglitz (1971)

5.1 Increasing Differences and FOSD

5.1.1 Comparative Statics

In this chapter, we consider the case where the parameter t is random and its realization is not learned until after the choice of some variable x. Examples include:

- The variable x could be the level of investment an individual makes in a financial asset, with t being the actual return on the asset (unknown at the time of investment).
- The variable x could be the amount of savings by an individual, with t being uncertain future income.
- The variable x could be the production level of a firm, with t being an uncertain parameter that affects market demand.

Suppose the distribution of t is affected by some additional parameter θ. That is, the parameter t has a cumulative distribution function $F_\theta(t)$, which itself depends on θ. There are several interpretations of this modeling assumption:

- The parameter θ could represent a signal is informative about t, with $F_\theta(t)$ representing the updated beliefs about the value of t following the signal realization θ.
- Alternatively, the parameter θ could represent the actions taken by some outside agent (exogenous from the perspective of our decision maker) that directly influence the distribution of t.

In this section, we will consider the case where increasing θ leads to a first-order stochastic dominance increase in the distribution F_θ. In other words, we will assume that $\theta' > \theta$ implies $F_{\theta'} \geq_{FOSD} F_\theta$. The examples given in Section 4.1 can be indexed by a parameter θ to form such families. For instance, consider again the examples of the uniform distribution from Example 4.5 and the discrete distribution from Example 4.6.

EXAMPLE 5.1 (UNIFORM DISTRIBUTION). Suppose $T = \mathbb{R}$, and suppose F_θ is the uniform distribution on the interval $[0, \theta]$ for $\theta > 0$. That is,

$$F_\theta(t) = \begin{cases} 0 & \text{if } t < 0 \\ t/\theta & \text{if } 0 \leq t \leq \theta \\ 1 & \text{if } \theta < t. \end{cases}$$

Then, $F_{\theta'} \geq_{FOSD} F_\theta$ whenever $\theta' > \theta$.

Example 5.2 (Binary discrete distributions). Suppose $T = \mathbb{R}$, and suppose that for each $\theta \in \mathbb{R}$ the distribution F_θ assigns some fixed probability $\alpha \in [0, 1]$ to the point θ and assigns probability $1 - \alpha$ to the point 0. Then, $F_{\theta'} \geq_{FOSD} F_\theta$ whenever $\theta' > \theta$.

In addition, there are many examples of established families of distributions that satisfy this ordering, such as:

- The normal distribution on $T = \mathbb{R}$ with mean θ for $\theta \in \mathbb{R}$.
- The exponential distribution on $T = \mathbb{R}_+$ with mean θ for $\theta > 0$.
- The Poisson distribution on $T = \mathbb{Z}_+$ with mean θ for $\theta > 0$.

We now provide our first comparative statics result for uncertain parameters t with distributions parameterized by θ.

Theorem 5.3. Suppose $X \subseteq \mathbb{R}^n$ is a lattice, $T \subseteq \mathbb{R}$, and $\Theta \subseteq \mathbb{R}$. Suppose $u : X \times T \to \mathbb{R}$ has increasing differences in $(x; t)$, and $\{F_\theta\}_{\theta \in \Theta}$ is a family of cumulative distribution functions of T such that $F_{\theta'} \geq_{FOSD} F_\theta$ if $\theta' > \theta$. Then the function $U : X \times \Theta \to \mathbb{R}$ defined by

$$U(x, \theta) = \int u(x, t)\, dF_\theta(t)$$

has increasing differences in $(x; \theta)$.[1]

Proof. Fix any $x' > x$. Define a function $g : T \to \mathbb{R}$ by

$$g(t) = u(x', t) - u(x, t).$$

Since u has increasing differences in $(x; t)$, the function g is nondecreasing. Fix any $\theta' > \theta$.

[1]Note that we are implicitly assuming in this result that $u(x, t)$ is integrable with respect to F_θ for each $x \in X$ and $\theta \in \Theta$, that is, the integral exists and is finite.

Since $F_{\theta'} \geq_{FOSD} F_\theta$, this implies

$$\begin{aligned} U(x',\theta') - U(x,\theta') &= \int g(t)\, dF_{\theta'}(t) \\ &\geq \int g(t)\, dF_\theta(t) = U(x',\theta) - U(x,\theta). \end{aligned}$$

Thus, U has increasing differences in $(x;\theta)$. □

In the case of a single-dimensional choice variable, i.e., $X \subseteq \mathbb{R}$, Theorem 5.3 gives us sufficient conditions for monotone comparative statics in θ. However, if $X \subseteq \mathbb{R}^n$ for $n > 1$, then our comparative statics results require additional complementarity between choice variables in the form of either supermodularity or quasisupermodularty of U in x. The following result shows that supermodularity of u in x gives us the same property of U.

Theorem 5.4. Suppose $X \subseteq \mathbb{R}^n$ is a lattice and $T \subseteq \mathbb{R}$. Suppose $u : X \times T \to \mathbb{R}$ is supermodular in x. Then for any cumulative distribution function F on T, the function $U : X \to \mathbb{R}$ defined by

$$U(x) = \int u(x,t)\, dF(t)$$

is supermodular in x.

REMARK 5.5. If the distribution is parameterized by θ, then this result obviously still holds. That is, if $u(x,t)$ is supermodular in x and $\{F_\theta\}_{\theta \in \Theta}$ is a family of cumulative distributions, then

$$U(x,\theta) = \int u(x,t)\, dF_\theta(t)$$

is supermodular in x. We omit the θ in Theorem 5.4 to emphasize that this parameter plays no role in the result.

Proof. Fix any $x, x' \in X$. Supermodularity of u in x implies that

$$u(x \wedge x', t) + u(x \vee x', t) \geq u(x,t) + u(x',t).$$

Since this is true for every t, integrating with respect to the distribution F_θ gives

$$\begin{aligned} U(x \wedge x') + U(x \vee x') &= \int \big[u(x \wedge x', t) + u(x \vee x', t)\big]\, dF(t) \\ &\geq \int \big[u(x,t) + u(x',t)\big]\, dF(t) \\ &= U(x) + U(x'). \end{aligned}$$

Thus U is supermodular in x. □

Corollary 5.6. Let $X \subseteq \mathbb{R}^n$ be a lattice, $T \subseteq \mathbb{R}$, and $\Theta \subseteq \mathbb{R}$. Suppose $u : X \times T \to \mathbb{R}$ is supermodular in x and has increasing differences in $(x;t)$, and suppose $\{F_\theta\}_{\theta\in\Theta}$ is a family of cumulative distribution functions such that $F_{\theta'} \geq_{FOSD} F_\theta$ if $\theta' > \theta$. If $U : X \times \Theta \to \mathbb{R}$ is defined by

$$U(x,\theta) = \int u(x,t)\, dF_\theta(t),$$

then $\operatorname{argmax}_{x\in X} U(x,\theta)$ is monotone nondecreasing in θ (in the strong set order).

Proof. Theorems 5.3 and 5.4 imply that U has increasing differences in $(x;\theta)$ and is supermodular in x. The monotonicity of the solution set then follows from our previous comparative statics result (Theorem 1.19). □

5.1.2 Applications

Example 5.7 (Monopolist facing uncertain demand). Suppose a monopoly faces uncertain demand and must make a production decision prior to learning the realized demand for its product. It does learn some information about demand prior to choosing its output. Formally, suppose the inverse demand function $P(q,t)$ depends on output and an unknown parameter t. The ex-post profit of the firm is therefore

$$\pi(q,t) = qP(q,t) - C(q),$$

where $C(q)$ is the cost function for the firm. Suppose the firm observes a signal θ that is informative about the parameter t. Specifically, suppose the cumulative distribution of t conditional on θ is F_θ. The ex-ante profit function for the firm is therefore[2]

$$\Pi(q,\theta) = \int \big[qP(q,t) - C(q)\big]\, dF_\theta(t).$$

If π has increasing differences in $(q;t)$ and $F_{\theta'} \geq_{FOSD} F_\theta$ for $\theta' > \theta$, then the firm's output $\operatorname{argmax}_{q\geq 0} \Pi(q,\theta)$ is nondecreasing in θ. Note that if π is differentiable, then increasing differences is equivalent to

$$\frac{\partial^2 \pi}{\partial q \partial t} = P_t(q,t) + qP_{qt}(q,t) \geq 0,$$

that is, marginal revenue is nondecreasing in t. For example, one special case where this holds is if $P(q,t) = \hat{P}(q) + t$.

[2] These formulas make the implicit assumption that the firm is fully committed to its production decision q before learning the realization of t. One could consider the alternative assumption that the firm chooses its inventory (production) $\bar{q}$ but is able to sell below inventory, $q < \bar{q}$, and dispose of the remaining units. However, our simplifying assumption that the firm chooses its exact output does not seem completely unreasonable. For example, selling less than inventory requires some commitment by the firm not to sell any remaining quantity at a lower price in a future period, which may not be realistic.

5.1.3 Limitations

While the comparative statics result from Corollary 5.6 is useful in many applications, it also has some limitations. The following example illustrates a case in which the assumptions of this result do not hold and we do not obtain the monotone comparative statics that we might have expected.

EXAMPLE 5.8 (INVESTMENT IN RISKY ASSET). Suppose an individual is allocating wealth w between a risk-free asset with gross return R^f and a risky asset with gross return $\tilde{R}$. If the individual invests α dollars in the risky asset (and the rest in the safe asset), then her future consumption will be

$$(w-\alpha)R^f + \alpha\tilde{R} = wR^f + \alpha(\tilde{R} - R^f).$$

Suppose the individual is an expected-utility maximizer with Bernoulli utility function $u(x)$. The indirect utility function for the individual as a function of α is therefore

$$U(\alpha) = \mathbb{E}\big[u(wR^f + \alpha(\tilde{R} - R^f))\big],$$

where the expectation is over the realization of the random return $\tilde{R}$. The objective of the individual is to choose α to maximize this value function.

One might conjecture that a first-order stochastic dominance shift in the distribution of the risky asset would lead to an increase in the investment α in the risky asset. Unfortunately, this is not correct as we now show. Make the following assumptions:

$$\begin{aligned}
&u(x) = \min\{x, 100\} \\
&w = 80 \\
&R^f = 1 \\
&\tilde{R} = \begin{cases} 2 & \text{with prob } p \\ 0 & \text{with prob } 1-p, \end{cases} \qquad\qquad p > 1/2.
\end{aligned}$$

It is useful to write the excess return of the risky asset:

$$\tilde{R} - R^f = \begin{cases} 1 & \text{with prob } p \\ -1 & \text{with prob } 1-p. \end{cases}$$

Then, it is not difficult to show that the individual's optimal investment is $\alpha^* = 20$, giving consumption

$$wR^f + \alpha^*(\tilde{R} - R^f) = \begin{cases} 100 & \text{with prob } p \\ 60 & \text{with prob } 1-p. \end{cases}$$

Suppose that the return on the risky asset changes and is now instead

$$\tilde{R} = \begin{cases} 3 & \text{with prob } p \\ 0 & \text{with prob } 1-p \end{cases}$$

which implies

$$\tilde{R} - R^f = \begin{cases} 2 & \text{with prob } p \\ -1 & \text{with prob } 1-p. \end{cases}$$

Notice that this is a first-order stochastic dominance increase in the distribution of $\tilde{R}$. However, it is easy to see that the individual's optimal investment now decreases to $\alpha^* = 10$, giving consumption

$$wR^f + \alpha^*(\tilde{R} - R^f) = \begin{cases} 100 & \text{with prob } p \\ 70 & \text{with prob } 1-p. \end{cases}$$

Note that the first-order stochastic dominance increase in the distribution of $\tilde{R}$ does not decrease the expected utility of the individual (think about why this must be true for any nondecreasing Bernoulli utility function) and in fact strictly increases her utility. However, it does decrease the amount the individual invests in the risky asset.

In light of this example, there are two possibilities for recovering a monotone comparative statics result. The first is to impose a stronger (more restrictive) ordering of distributions. We will pursue this approach in the next section. The other possibility is to explore more restrictive classes of expected utility functions for which a first-order stochastic dominance shift in the distribution of a risky asset leads to an increase in investment. The restrictions on utility functions needed for such results are fairly strong, and we will not pursue them here.[3]

5.2 Single Crossing and Log-Supermodularity

5.2.1 Log-Supermodularity and the MLR Property

Definition 5.9. Let $X \subseteq \mathbb{R}^n$ be a lattice. A function $h : X \to \mathbb{R}_+$ is **log-supermodular (log-SM)** if for all $x, x' \in X$,

$$h(x \wedge x') \cdot h(x \vee x') \geq h(x) \cdot h(x').$$

To understand this terminology, note that if h is strictly positive, then it is log-SM if and only if $\log(h(x))$ is a supermodular function.

[3] One sufficient condition is that relative risk aversion be less than unity. The interested reader can consult Gollier (2001, page 61).

In this section, we will typically be interested in a log-supermodular function of two parameters. Suppose $T \subseteq \mathbb{R}$ and $\Theta \subseteq \mathbb{R}$. In this case, a function $h : T \times \Theta \to \mathbb{R}$ is log-SM if for all $t' > t$ and $\theta_H > \theta_L$,

$$h(t, \theta_L) \cdot h(t', \theta_H) \geq h(t, \theta_H) \cdot h(t', \theta_L).$$

If h is strictly positive, this can be rewritten as

$$\frac{h(t', \theta_H)}{h(t', \theta_L)} \geq \frac{h(t, \theta_H)}{h(t, \theta_L)},$$

that is, $h(t, \theta_H)/h(t, \theta_L)$ is nondecreasing in t for $\theta_H > \theta_L$.

In the case where $h(\cdot, \theta)$ is a probability density function for each θ, h is log-supermodular if and only if the family of densities $\{h(\cdot, \theta)\}_{\theta \in \Theta}$ has the **monotone likelihood ratio property**, that is, the family respects the monotone likelihood ratio order with respect to increases in θ. Several families of probability distributions mentioned previously exhibit this property, including:

- The uniform distribution on $[0, \theta]$ for $\theta > 0$.
- The normal distribution with mean θ for $\theta \in \mathbb{R}$.
- The exponential distribution with mean θ for $\theta > 0$.
- The Poisson distribution with mean θ for $\theta > 0$.

However, since the MLR order is more restrictive than the FOSD dominance order, not all families of distributions that are ordered according to first-order stochastic dominance have the MLR property. For instance, the discrete distribution described in Example 5.2 does *not* have the monotone likelihood ratio property (see Example 4.13 for a related discussion).

5.2.2 Comparative Statics

The results in this section appear in various forms in a number of papers and books, including Karlin (1968), Shannon (1995), Gollier (2001), and Athey (2002). In particular, Theorems 5.10 and 5.12 appear in a slightly stronger form in Athey (2002, Theorem 2 and Lemma 5).

Theorem 5.10. Suppose $X \subseteq \mathbb{R}^n$ is a lattice, $T \subseteq \mathbb{R}$, and $\Theta \subseteq \mathbb{R}$. Suppose $u : X \times T \to \mathbb{R}$ has the single crossing property in $(x; t)$ and $h : T \times \Theta \to \mathbb{R}_+$ is log-supermodular. Then:

(1) If T is an interval and $h(t, \theta) > 0$ for all $t \in T$ and $\theta \in \Theta$, then $U : X \times \Theta \to \mathbb{R}$ defined by

$$U(x, \theta) = \int u(x, t) h(t, \theta)\, dt$$

has the single crossing property in $(x; \theta)$.

(2) If there exists a finite set $T^* \subseteq T$ such that $h(t,\theta) > 0$ for all $t \in T^*$ and $\theta \in \Theta$, then $U : X \times \Theta \to \mathbb{R}$ defined by

$$U(x,\theta) = \sum_{t \in T^*} u(x,t)h(t,\theta)$$

has the single crossing property in $(x;\theta)$.[4]

In particular, when $h(\cdot,\theta)$ is a probability density function, the two claims in this theorem correspond to the two cases we considered when defining the monotone likelihood ratio order in Section 4.2: continuous and discrete distributions. The proof of Theorem 5.10 will be based on the following definition and result, which we will also use directly in some applications.

Definition 5.11. Suppose $T \subseteq \mathbb{R}$ and $g : T \to \mathbb{R}$. We say g has the **(single variable) single crossing property** if for any $t' > t$,

$$\begin{aligned} g(t) \geq 0 &\implies g(t') \geq 0, \quad \text{and} \\ g(t) > 0 &\implies g(t') > 0. \end{aligned}$$

Note the connection between this definition of single crossing and our prior definition. A function $u(x,t)$ satisfies the (multivariate) single crossing property in $(x;t)$ if and only if for any $x' > x$ the function

$$g(t) = u(x',t) - u(x,t)$$

satisfies the (single variable) single crossing property.

Theorem 5.12. Suppose $T \subseteq \mathbb{R}$ and $\Theta \subseteq \mathbb{R}$, and suppose $g : T \to \mathbb{R}$ has the (single variable) single crossing property[5] and $h : T \times \Theta \to \mathbb{R}_+$ is log-supermodular. Then:

(1) If T is an interval and $h(t,\theta) > 0$ for all $t \in T$ and $\theta \in \Theta$, then

$$\varphi(\theta) = \int g(t)h(t,\theta)\,dt$$

has the (single variable) single crossing property.

[4] This result also holds for countably infinite $T^* \subseteq T$, provided T^* is closed and the infinite series is absolutely convergent (meaning the order in which the elements of T^* are enumerated does not affect the value of the sum).

[5] In fact, as will be evident from the proof, a weaker condition on g is sufficient for the conclusion of this theorem. It suffices to assume that g satisfies the *weak* single crossing property: $g(t) > 0$ implies $g(t') \geq 0$ for all $t' > t$. Weak single crossing is equivalent to the requirement that either (i) $g(t) > 0$ for all t, (ii) $g(t) < 0$ for all t, or (iii) there exists t_0 such that $g(t) \leq 0$ for $t < t_0$ and $g(t) \geq 0$ for $t > t_0$.

(2) If there exists a finite set $T^* \subseteq T$ such that $h(t,\theta) > 0$ for all $t \in T^*$ and $\theta \in \Theta$, then

$$\varphi(\theta) = \sum_{t \in T^*} g(t)h(t,\theta)$$

has the (single variable) single crossing property.

Proof. We first prove claim 1. Assume that T is an interval and that $h(t,\theta) > 0$ for all $t \in T$ and $\theta \in \Theta$. Since g satisfies the single crossing property, there are three cases to consider:

Case 1 — $g(t) > 0$ for all t: This implies $\varphi(\theta) > 0$ for all θ, so the single crossing property is trivially satisfied.

Case 2 — $g(t) < 0$ for all t: This implies $\varphi(\theta) < 0$ for all θ, so the single crossing property is again trivially satisfied.

Case 3 — There exists t_0 such that $g(t) \leq 0$ for all $t < t_0$ and $g(t) \geq 0$ for all $t > t_0$: Fix any $\theta_H > \theta_L$. By the definition of log-SM, we have

$$\begin{aligned} t > t_0 &\implies h(t,\theta_H)h(t_0,\theta_L) \geq h(t,\theta_L)h(t_0,\theta_H) \\ t < t_0 &\implies h(t,\theta_H)h(t_0,\theta_L) \leq h(t,\theta_L)h(t_0,\theta_H). \end{aligned}$$

Rearranging these inequalities and using the definition of t_0, we have

$$\begin{aligned} t > t_0 &\implies h(t,\theta_H) \geq h(t,\theta_L)\frac{h(t_0,\theta_H)}{h(t_0,\theta_L)} \quad \text{and} \quad g(t) \geq 0 \\ t = t_0 &\implies h(t,\theta_H) = h(t,\theta_L)\frac{h(t_0,\theta_H)}{h(t_0,\theta_L)} \\ t < t_0 &\implies h(t,\theta_H) \leq h(t,\theta_L)\frac{h(t_0,\theta_H)}{h(t_0,\theta_L)} \quad \text{and} \quad g(t) \leq 0. \end{aligned}$$

Therefore, for any t,

$$g(t)h(t,\theta_H) \geq g(t)h(t,\theta_L)\frac{h(t_0,\theta_H)}{h(t_0,\theta_L)}.$$

Integrating with respect to t gives

$$\begin{aligned} \varphi(\theta_H) &= \int g(t)h(t,\theta_H)\,dt \\ &\geq \frac{h(t_0,\theta_H)}{h(t_0,\theta_L)} \int g(t)h(t,\theta_L)\,dt = \frac{h(t_0,\theta_H)}{h(t_0,\theta_L)}\varphi(\theta_L). \end{aligned}$$

Since h is strictly positive, it follows that

$$\begin{aligned} \varphi(\theta_L) \geq 0 &\implies \varphi(\theta_H) \geq 0 \\ \varphi(\theta_L) > 0 &\implies \varphi(\theta_H) > 0. \end{aligned}$$

Thus φ has the single crossing property.

We now prove claim 2. Assume that $T^* \subseteq T$ is finite and that $h(t, \theta) > 0$ for all $t \in T^*$ and $\theta \in \Theta$. Consider the same three cases outlined in the proof of claim 1. In case 1 ($g(t) > 0$ for all t) and case 2 ($g(t) < 0$ for all t), single crossing of φ is again trivially satisfied. Consider case 3, so there exists $t_0 \in T$ such that $g(t) \leq 0$ for all $t < t_0$ and $g(t) \geq 0$ for all $t > t_0$. Let t_0^* be either the smallest element of T^* that is weakly larger than t_0 or the largest element of T^* that is weakly smaller than t_0. Then, we have $h(t_0^*, \theta) > 0$ for all $\theta \in \Theta$, and the remainder of the proof proceeds exactly as in the proof of claim 1, but with t_0^* in the place of t_0 and with sums in the place of integrals. $\square$

Proof of Theorem 5.10. We will prove claim 1. The proof of claim 2 is analogous. Fix any $x' > x$, and define $g : T \to \mathbb{R}$ by

$$g(t) = u(x', t) - u(x, t).$$

Since u satisfies the single crossing property in $(x; t)$, the function g has the (single variable) single crossing property. By Theorem 5.12, the function

$$\varphi(\theta) = \int g(t)h(t, \theta)\, dt = U(x', \theta) - U(x, \theta)$$

has the (single variable) single crossing property. Since this is true for any $x' > x$, the function U has the single crossing property in $(x; \theta)$. $\square$

Corollary 5.13. Suppose $X \subseteq \mathbb{R}^n$ is a lattice, $T \subseteq \mathbb{R}$ is an interval, and $\Theta \subseteq \mathbb{R}$. Suppose $u : X \times T \to \mathbb{R}$ is supermodular in x and has the single crossing property in $(x; t)$, and suppose $h : T \times \Theta \to \mathbb{R}$ is log-supermodular and strictly positive. If $U : X \to \mathbb{R}$ is defined by

$$U(x, \theta) = \int u(x, t)h(t, \theta)\, dt,$$

then $\operatorname{argmax}_{x \in X} U(x, \theta)$ is monotone nondecreasing in θ (in the strong set order).

Proof. Theorems 5.4 and 5.10 imply that U is supermodular in x and has the single crossing property in $(x; \theta)$. The monotonicity of the solution set then follows from our previous ordinal comparative statics result (Theorem 2.12). $\square$

REMARK 5.14. Athey (2002, Lemma 4 and Theorem 1) also shows that if both $u(x, t)$ and $h(t, \theta)$ are log-SM, then $U(x, \theta) = \int u(x, t)h(t, \theta)dt$ is log-SM. Note that log-SM of u (respectively, U) implies that u (respectively, U) satisfies the single crossing property (recall that any monotone transformation of a supermodular function satisfies the single crossing property). These results therefore place stronger restrictions on u in order to obtain a stronger condition on U than is needed for monotone comparative statics. However, the advantage of these results is that they can apply to $T \subseteq \mathbb{R}^m$ for $m \geq 2$, whereas the results in this section restrict to $T \subseteq \mathbb{R}$.

5.2.3 Applications

Our first application revisits our previous investment example, but now considers an MLR change in the distribution of the risky asset.

EXAMPLE 5.15 (INVESTMENT PROBLEM WITH MLR CHANGE IN RISK). The setup is similar to in Example 5.8. An individual is allocating wealth w between a risk-free asset with gross return R^f and a risky asset with gross return $\tilde{R}$. If the individual invests α dollars in the risky asset (and the rest in the safe asset), then her future consumption will be

$$(w-\alpha)R^f + \alpha\tilde{R} = wR^f + \alpha(\tilde{R} - R^f).$$

Suppose $\tilde{R}$ takes values on some interval $T \subseteq \mathbb{R}$, and suppose the distribution of $\tilde{R}$ has a density function $f_\theta(t)$ that is strictly positive on this interval. In addition, suppose $\theta' > \theta$ implies $f_{\theta'} \geq_{MLR} f_\theta$. This assumption is in contrast to Example 5.8, where we considered FOSD rather than MLR changes in the distribution of the risky asset.

Suppose the individual is an expected-utility maximizer with a strictly increasing Bernoulli utility function $u(x)$. The indirect utility function for the individual as a function of the investment level α and the parameter θ is therefore

$$U(\alpha, \theta) = \int u(wR^f + \alpha(t - R^f))f_\theta(t)\,dt.$$

The objective of the individual is to choose α to maximize this value function. Let

$$v(\alpha, t) = u(wR^f + \alpha(t - R^f)),$$

so the value function can be written as

$$U(\alpha, \theta) = \int v(\alpha, t)f_\theta(t)\,dt.$$

Notice that $v(\alpha, t)$ has the single-crossing property in $(\alpha; t)$. To see this, fix any $\alpha' > \alpha$ and

$t' > t$, and note that

$$\begin{aligned}
v(\alpha', t) \geq v(\alpha, t) &\iff u(wR^f + \alpha'(t - R^f)) \geq u(wR^f + \alpha(t - R^f)) \\
&\iff \alpha'(t - R^f) \geq \alpha(t - R^f) \qquad \text{(since } u \text{ strictly increasing)} \\
&\iff (\alpha' - \alpha)(t - R^f) \geq 0 \\
&\iff t - R^f \geq 0 \qquad \text{(since } \alpha' > \alpha\text{)} \\
&\implies t' - R^f > 0 \qquad \text{(since } t' > t\text{)} \\
&\iff u(wR^f + \alpha'(t' - R^f)) > u(wR^f + \alpha(t' - R^f)) \\
&\iff v(\alpha', t') > v(\alpha, t').
\end{aligned}$$

Therefore, since $v(\alpha, t)$ has the single-crossing property in $(\alpha; t)$ and $f_\theta(t)$ is log-supermodular, the function $U(\alpha, \theta)$ has the single-crossing property in $(\alpha; \theta)$ by Theorem 5.10. In particular, this implies that the set of maximizing α are nondecreasing in θ in the strong set order (Theorem 2.2). In the special case where u is strictly concave, there is a unique solution $\alpha^*(\theta)$ which is nondecreasing in θ.

The following result will be useful for our second application. The proof is straightforward, so we omit it.

Lemma 5.16. Suppose $X \subseteq \mathbb{R}$ is a closed interval, $\Theta \subseteq \mathbb{R}$, and $U : X \times \Theta \to \mathbb{R}$ is twice differentiable with respect to x and satisfies $U_{xx} < 0$. Suppose that the problem of choosing x to maximize this objective function has a solution for every θ, denoted $x^*(\theta) \in \operatorname{argmax}_{x \in X} U(x, \theta)$. (Note that strict concavity implies that this solution is unique.) Fix any $\theta' > \theta$. If U satisfies

$$U_x(x^*(\theta), \theta) \geq 0 \implies U_x(x^*(\theta), \theta') \geq 0,$$

then $x^*(\theta') \geq x^*(\theta)$.

In particular, this lemma implies that a sufficient condition for $x^*(\theta)$ to be nondecreasing in θ is that for all $x \in X$ and for all $\theta' > \theta$,

$$U_x(x, \theta) \geq 0 \implies U_x(x, \theta') \geq 0.$$

However, as the lemma shows, we actually only need this condition to be satisfied at $x = x^*(\theta)$. In addition, the inequality on the left must be satisfied with equality if $x^*(\theta)$ is an interior solution, as we can only have $U_x(x^*(\theta), \theta) < 0$ if $x^*(\theta)$ is the left boundary point of X and $U_x(x^*(\theta), \theta) > 0$ if $x^*(\theta)$ is the right boundary point of X.

Example 5.17 (Investment problem with change in risk aversion). The setup is the same as in the previous example. However, suppose that the distribution of the risky asset $\tilde{R}$ is fixed and has a density function $f(t)$. We now instead explore the impact of changes in risk aversion. Specifically, assume that individual 1 is more risk averse than individual 2. For this example, we assume the the Bernoulli utility functions u_i of these individuals are twice differentiable with $u_i' > 0$ and $u_i'' < 0$. The value function is

$$U(\alpha, i) = \int u_i(wR^f + \alpha(t - R^f))f(t)\,dt.$$

Note that

$$U_\alpha(\alpha, i) = \int u_i'(wR^f + \alpha(t - R^f))(t - R^f)f(t)\,dt,$$

$$U_{\alpha\alpha}(\alpha, i) = \int u_i''(wR^f + \alpha(t - R^f))(t - R^f)^2 f(t)\,dt < 0.$$

Fix any α, and let

$$g(t) = (t - R^f)f(t)$$
$$h(t, i) = u_i'(wR^f + \alpha(t - R^f))$$

Clearly, g has the (single variable) single crossing property. The only challenge is to show that h is log-SM. We will show that this is true provided $\alpha \geq 0$. Since $h > 0$, we have

$$\begin{aligned} h(t,i) \text{ is log-SM} &\iff \log(h(t,i)) \text{ is SM} \\ &\iff \frac{\partial}{\partial t}\log(h(t,1)) \leq \frac{\partial}{\partial t}\log(h(t,2)) \\ &\iff \frac{h_t(t,1)}{h(t,1)} \leq \frac{h_t(t,2)}{h(t,2)} \\ &\iff \frac{u_1''(wR^f + \alpha(t - R^f))}{u_1'(wR^f + \alpha(t - R^f))}\alpha \leq \frac{u_2''(wR^f + \alpha(t - R^f))}{u_2'(wR^f + \alpha(t - R^f))}\alpha. \end{aligned}$$

If $\alpha = 0$, then this condition is trivially satisfied. If $\alpha > 0$, then this condition is equivalent to

$$A_1(wR^f + \alpha(t - R^f)) \geq A_2(wR^f + \alpha(t - R^f)),$$

where $A_i(x)$ denotes the Arrow-Pratt coefficient of absolute risk aversion for individual i at x (this inequality holds since individual 1 is more risk averse than individual 2). Therefore, holding $\alpha \geq 0$ fixed,

$$\varphi(i) \equiv \int g(t)h(t,i)\,dt = U_\alpha(\alpha, i)$$

has the (single variable) single crossing property by Theorem 5.12. In other words, for any $\alpha \geq 0$,

$$U_\alpha(\alpha, 1) \geq 0 \implies U_\alpha(\alpha, 2) \geq 0.$$

Hence, by Lemma 5.16, if $\alpha_1^* \geq 0$ then $\alpha_2^* \geq \alpha_1^*$.

Interestingly, if $\alpha \leq 0$, then the arguments above are reversed and we have

$$U_\alpha(\alpha, 1) \leq 0 \implies U_\alpha(\alpha, 2) \leq 0.$$

Thus, if $\alpha_1^* \le 0$ (so individual 1 finds it optimal to short sell the asset), then $\alpha_2^* \le \alpha_1^*$ (individual 2 finds a weakly larger short sale optimal). This has a natural economic interpretation. A larger short sale implies greater exposure to risk, and individual 2 is less risk averse and therefore more willing to bear such risk. Combining these two cases, we see that $\alpha_1^* = 0$ if and only if $\alpha_2^* = 0$ (this can also be proved directly by using first-order conditions to show that the sign of α_i^* is the same as the sign of $\mathbb{E}[\tilde{R} - R^f]$).

5.3 Exercises

5.1 Suppose $X \subseteq \mathbb{R}^n$ is a lattice, and suppose that $f : X \to \mathbb{R}_+$ and $g : X \to \mathbb{R}_+$ are log-supermodular. Is the function $h : X \to \mathbb{R}_+$ defined by $h(x) = f(x)g(x)$ log-supermodular? Prove or provide a counterexample.

5.2 Suppose a firm produces a product using capital together with the labor of a single employee. Different potential employees have different levels of productivity, and the firm's output depends on the amount of capital used by the firm and the productivity of the employee it hires (i.e., units of effective labor). The firm conducts interviews of applicants and is able to fully learn the productivity of each applicant that it interviews; it hires the best among the pool of applicants. Formally, suppose the firm's production function is

$$f(k, x) = k^\alpha x^\beta, \quad \text{where } \alpha, \beta > 0 \text{ and } \alpha + \beta \le 1.$$

The variable $k \ge 0$ is the number of units of capital, and $x \ge 0$ is the productivity of the employee of the firm. The input and output markets are competitive, with price p for each unit of output, wage w for the employee (independent of the productivity of the employee), and price r for each unit of capital. If the firm operates, its profit is therefore

$$\pi(k, x) = pf(k, x) - w - rk.$$

The productivity of each applicant is independently and identically distributed (iid) according to $G(x)$. The timing is as follows:

1. First, the firm learns how many applicants it has for the position. Let $t \in \mathbb{N}$ denote the number of applicants. (Treat t as an exogenous parameter.)
2. Before learning the productivity of the applicants, the firm must decide how many units of capital to purchase.
3. The firm then interviews the applicants, learns the productivity of each, and chooses whether to hire the applicant with the highest productivity or to hire no one and shut down it's production operation (in this case, the cost of capital is a loss for the firm but the wage cost is avoided).

Answer the following:

(a) Assume first that the firm must operate. That is, it does not have the option to shut down, and it will therefore hire the most qualified applicant. Can you say anything about how the choice of capital k varies with the number of applicants t? That is, is there a monotonic (increasing or decreasing) relationship between them? If yes, prove it. If not, provide a counterexample.

(b) Assume now that the firm has the option to shut down if the best applicant is not sufficiently productive. Does this change your answer to part (a)? Provide a proof or a counterexample to support your answer.

(c) Assume again that the firm must operate and does not have the option to shut down. Is it possible to find an example of a production function $f(k, x)$ that is nondecreasing in k and x, but for which the optimal choice of capital is nonincreasing in the number of applicants t?

5.3 In Exercise 4.5, we explored the concepts of decreasing absolute risk aversion (DARA) and decreasing relative risk aversion (DRRA). In this problem, we study the implications of these concepts for a portfolio choice problem. Consider an investment problem as in the examples from the chapter, with a riskless asset with gross return R^f and a risky asset with gross return $\tilde{R}$. The individual chooses an investment $\alpha \in \mathbb{R}$ in the risky asset, which yields a stochastic consumption of

$$(w - \alpha)R^f + \alpha\tilde{R} = wR^f + \alpha(\tilde{R} - R^f).$$

Suppose $\mathbb{E}[\tilde{R}] > R^f$. You may also assume that the distribution of $\tilde{R}$ has a density function that is strictly positive on its support, and you may also assume a solution to the portfolio problem exists. The individual has a Bernoulli utility function u that is differentiable, strictly increasing, and strictly concave. Answer the following: (I am not expecting you to prove anything from first principles. The goal is for you to see how to apply the example from this chapter in a clever way that allows you to address the issues in this problem.)

(a) Suppose that the individual exhibits decreasing absolute risk aversion (DARA). Can we say anything about how the optimal amount that the individual will invest in the risky asset, α, varies with initial wealth, w? Prove any relationship that you claim.

(b) Suppose that the individual exhibits decreasing relative risk aversion (DRRA). Can we say anything about how the optimal fraction of wealth that the individual will invest in the risky asset, α/w, varies with initial wealth, w? Prove any relationship that you claim.

5.4 Suppose a monopoly faces uncertain demand and must make a production decision prior to learning the realized demand for its product. It does learn some information about demand prior to choosing its output. Formally, suppose the inverse demand function $P(q, t)$ depends on output and an unknown parameter t. The ex-post profit of the firm is therefore

$$\pi(q, t) = qP(q, t) - c(q),$$

where $c(q)$ is the cost function for the firm. Suppose the firm observes a signal θ that is informative about the parameter t. Specifically, suppose the cumulative distribution of t conditional on θ is F_θ. The ex-ante profit function for the firm is therefore (assume that the firm is fully committed to its production decision q before learning the realization of t):

$$\Pi(q, \theta) = \int \big[qP(q, t) - c(q)\big]\, dF_\theta(t).$$

Suppose $F_{\theta'} \geq_{MPRR} F_\theta$ for $\theta' > \theta$. Suppose the inverse demand function takes the form $P(q, t) = g(q) + h(t)$. What conditions on g and h imply that the firm's output $\operatorname{argmax}_{q \geq 0} \Pi(q, \theta)$ is nondecreasing in θ?

5.5 Consider a monopoly choosing its price to maximize profit. Suppose the demand function depends on the price p and a parameter t, and takes the form

$$D(p, t) = (a - bp + t)^\alpha.$$

Assume that $a, b, \alpha > 0$ are all positive constants. The firm has a constant marginal cost $c > 0$ of production (assume $a/b > c$), and therefore has a profit function

$$\pi(p, t) = (p - c)D(p, t).$$

Assume that the parameter t can take any nonnegative value, $t \geq 0$, and assume that the firm can choose any price $p \in [c, \bar{p}]$, where $\bar{p} = a/b$. (We could permit $p < c$, but this is clearly never optimal, so we will make the technically convenient assumption that p must be greater than c.)

(a) For which values of $\alpha > 0$ does π have increasing differences in $(p; t)$?

(b) For which values of $\alpha > 0$ does π have the single crossing property in $(p; t)$?

(c) Suppose that the monopoly faces uncertain demand, which we will model as uncertainty about the parameter t. Assume that the price p must be chosen prior to learning the realization of t. (For example, if the monopoly publicly advertises its price p in advance of learning demand and cannot changed it after observing the realization of demand, then this assumption would be reasonable.) For a cumulative distribution F of the parameter t, the ex-ante profit of the firm is therefore

$$\Pi(p, F) = \int \pi(p, t)\, dF(t).$$

Consider first the benchmark case of $\alpha = 1$. In this case, find an *explicit solution* for the ex-ante profit-maximizing price p as a function of the distribution F (as well as the constants a, b, c). Does a first-order stochastic dominance increase in the distribution of t lead to a (weakly) higher optimal price p?

(d) Consider again the ex-ante profit function $\Pi(p, F)$ defined in the previous part of the problem. For $\alpha \neq 1$, it may not be possible to find an explicit solution for the optimal p as a function of F, but we still may be able to obtain comparative statics in some cases. For which values of $\alpha > 0$ does a first-order stochastic dominance shift in the distribution of t lead to an increase in the optimal price p? That is, for which $\alpha > 0$ does $F \geq_{FOSD} G$ imply that $\operatorname{argmax}_{p\in[c,\bar{p}]} \Pi(p, F)$ is greater than $\operatorname{argmax}_{p\in[c,\bar{p}]} \Pi(p, G)$ in the strong set order? (Note: I am asking you to identify the set of values of α for which our theorems from class can be applied to this problem to obtain monotone comparative statics, and to carefully explain what the relevant theorems say and why they can be applied. I am not requiring that you provide counterexamples to show that we fail to have monotone comparative statics for other values of α.)

(e) Consider again the ex-ante profit function $\Pi(p, F)$. For which values of $\alpha > 0$ does a monotone likelihood ratio shift in the distribution of t lead to an increase in the optimal price p? That is, for which $\alpha > 0$ does $F \geq_{MLR} G$ imply that $\operatorname{argmax}_{p\in[c,\bar{p}]} \Pi(p, F)$ is greater than $\operatorname{argmax}_{p\in[c,\bar{p}]} \Pi(p, G)$ in the strong set order? You can assume that F and G are continuously distributed and have density functions that are strictly positive on $\mathbb{R}_+$. (Note: Again, I am asking you to identify the set of values of α for which our theorems from class can be applied to this problem to obtain monotone comparative statics. I am not requiring that you provide counterexamples to show that we fail to have monotone comparative statics for other values of α.)

5.6 Consider a two-period consumption-savings problem. Suppose the individual has the following utility function for $(x, y) \in \mathbb{R}^2_+$:

$$U(x, y) = u(x) + v(y),$$

where u and v are nondecreasing and twice continuously differentiable functions. Assume the gross interest rate is normalized to unity, so the intertemporal budget constraint is simply $x + y = w$. Suppose wealth w is random (e.g., due to uncertain future income) and the individual chooses first-period consumption x before learning the realization of w. Thus the consumer's maximum expected utility when w has distribution F is

$$V(F) = \max_{x\geq 0} \int \Big(u(x) + v(w - x)\Big)\, dF(w).$$

Answer the following. In each part below, you are not required to find necessary conditions, but finding a weaker sufficient condition and avoiding unnecessary conditions will earn more points. (For example, saying u and v are constant may be correct for some parts, but will earn zero points.)

(a) Find a sufficient condition on the shape of u and/or v so that the individual prefers FOSD increases in the distribution of w (that is, $F \geq_{FOSD} G \Rightarrow V(F) \geq V(G)$).

(b) Restrict attention for this part to degenerate lotteries over wealth (i.e., deterministic w). Under what, if any, conditions on u and/or v is the optimal x nondecreasing (in the strong set order) in w? (In other words, when is demand for x normal?)

(c) Returning to the case of uncertain wealth, under what, if any, conditions on u and/or v is the optimal x nondecreasing (in the strong set order) when there is a FOSD increase in the distribution of w?

(d) Under what, if any, conditions on u and/or v is the optimal x nondecreasing (in the strong set order) when there is a mean-preserving reduction in risk (MPRR) in the distribution of w?

5.7 Let $N = \{1, \ldots, n\}$ denote a set of players in a simultaneous move game. Suppose each player has a strategy space $S_i = [0, 1]$, and suppose $T = [0, 1]$ is a set of parameters. The utility of player i is given by

$$u_i(s_i, s_{-i}, t) = (1 - \alpha)\phi(s_i - t) + \alpha \sum_{j \neq i} \phi(s_i - s_j).$$

In this equation, $\alpha \in (0, 1)$ is a fixed constant, and $\phi : \mathbb{R} \to \mathbb{R}$ is twice continuously differentiable with $\phi'' < 0$ and $\phi(0) = \phi'(0) = 0$. In words, ϕ is a strictly concave function that is everywhere below zero ($\phi(x) \leq 0$) and only attains the value zero at $x = 0$. (You might recognize that the game we have just defined is a version of famous the "beauty contest" game of Keynes.)

(a) Suppose the value of the parameter t is common knowledge. Based only on the assumptions given so far, what can be said about how the set of Nash equilibria of this game change in response to a change in t? Be precise, and prove any claim that you make. If you say that nothing can be determined based on the assumptions given so far, what additional assumptions would allow you to say something definitive about how the set of Nash equilibria change with t?

(b) Suppose instead that the parameter t is uncertain for the players at the time when they are choosing their strategies. Suppose the players have a common prior (cumulative distribution function) G over t. If the prior increases with respect to the first-order stochastic dominance order, that is, the prior changes from G to F where $F \geq_{FOSD} G$, then what can be said about how how the set of Nash equilibria of this game change?[6] Be precise, and prove any claim that you make. If you say that a definitive prediction cannot be made based on the assumptions given so far, what additional assumptions would allow you to say something definitive about how the set of NE change in response to a FOSD increase in the distribution of t?

[6] Intuitively, such a change in common beliefs could be the result of a public signal (meaning the signal is observable to all of the players) that makes the players more optimistic about the distribution of t (in the sense of FOSD). However, the exact cause of the change in beliefs from G to F is not important for solving this problem.

(c) As in the previous part of the problem, suppose again that the parameter t is uncertain for the players at the time when they are choosing their strategies and that the players have a common prior G over t. If the prior increases with respect to the second-order stochastic dominance order, that is, the prior changes from G to F where $F \geq_{SOSD} G$, then what can be said about how the set of Nash equilibria of this game change? Be precise, and prove any claim that you make. If you say that a definitive prediction cannot be made based on the assumptions given so far, what additional assumptions would allow you to say something definitive about how the set of NE change in response to a SOSD increase in the distribution of t?

5.8 This problem concerns a variation of the Le Chatelier Principle that we studied previously. We now consider a decision maker who is forward looking and anticipates that a change in the parameter value might occur (i.e., the exact change to the parameter is unknown but the decision maker has beliefs about its possible future values). Formally, consider the problem of selecting two variables x and y optimally for a given payoff relevant parameter t. Consider a multi-period problem where t may change over time. The variable x can respond immediately to changes in t. However, the variable y takes time to adjust and must be chosen one period in advance, which means that the value of t in a given period may not yet be known at the time when y must be chosen.

Specifically, suppose $X \subseteq \mathbb{R}$ and $Y \subseteq \mathbb{R}$ are compact, $T \subseteq \mathbb{R}$, and $f : X \times Y \times T \to \mathbb{R}$ is continuous and supermodular in (x, y) and has increasing differences in $(x, y; t)$. There are three time periods, $\tau = 0, 1, 2$ (note the use of τ to denote time periods, to avoid confusion with the parameter t). Suppose the that value of the parameter t is t_0 in period 0. In period 1, the value of the parameter changes to t_1 and it remains at t_1 in period 2. Thus, period 1 is the only period in which the parameter value might change (it is also possible that $t_1 = t_0$, in which case there is no change in the parameter value). The objective of the agent is to maximize

$$f(x_0, y_0, t_0) + f(x_1, y_1, t_1) + f(x_2, y_2, t_1).$$

However, the difficulty faced by the agent is that y_1 must be chosen in one period in advance (in period 0) before the realized value of the parameter t_1 is learned. Specifically, the timeline of information available to the agent and the timing of her decisions is as follows:

Period 0:

- The parameter value t_0 in period 0 is known well in advance, so the agent can choose both x_0 and y_0 optimally given this parameter value (y_0 needs to be chosen in advance of period 0, but t_0 was also known in advance, the agent was able to chose y_0 optimally for this period).
- The agent knows in period 0 that the distribution of t_1 is given by the cdf F,

and in period 0 she must select the next-period value y_1 of y using only this information.

Period 1:

- The agent learns the realization of t_1.
- She chooses x_1 optimally given this realized parameter value t_1 and the value $y = y_1$ that she previously selected in period 0.
- The agent chooses the next-period value y_2 of y for period 2.

Period 2:

- The parameter remains at t_1.
- The agent chooses the value x_2 of x.

The solutions to several optimization problems will be useful for describing these variables:

$$\begin{aligned}
(x^{LR}(t), y^{LR}(t)) &\in \underset{(x,y)\in X\times Y}{\operatorname{argmax}} \ f(x,y,t), \\
x^{SR}(y,t) &\in \underset{x\in X}{\operatorname{argmax}} \, f(x,y,t), \\
y^{EX}(F) &\in \underset{y\in Y}{\operatorname{argmax}} \int f(x^{SR}(y,t),y,t)\, dF(t).
\end{aligned}$$

For simplicity, you can assume throughout this problem that if there are multiple solutions to any of these maximization problems, the agent chooses the solution with the largest values of x and y. The notation "LR" stands for long run, "SR" stands for short run, and "EX" stands for expectation. The timeline of information and decisions given above implies that:[7]

$$\begin{array}{lll}
y_0 = y^{LR}(t_0) & y_1 = y^{EX}(F) & y_2 = y^{LR}(t_1) \\
x_0 = x^{LR}(t_0) & x_1 = x^{SR}(y_1, t_1) & x_2 = x^{LR}(t_1)
\end{array}$$

Answer the following:

(a) Define a function $g : Y \times T \to \mathbb{R}$ by $g(y,t) = f(x^{SR}(y,t),y,t)$.[8] Under the assumptions given above that f is supermodular in (x,y) and has increasing differences in $(x,y;t)$, prove that g has increasing differences in $(y;t)$. (Hint: If f

[7] Note that this is essentially the same setup that we used to study the Le Chatelier Principle in class, but with one important difference: In this problem, we will allow the decision maker to have beliefs F in period 0 about what the value t_1 of the parameter will be in period 1, so she can select y_1 in anticipation of these possible future changes to t. In class, we assumed that $y_1 = y^{LR}(t_0)$, which could be interpreted as the agent having the period 0 belief that with probably 1 the parameter value would remain at the value t_0 in period 1. Thus, when we studied the Le Chatelier Principle in class, we implicitly assumed that the change in parameter was a surprise (or probability zero event) to the agent.

[8] Thus $g(y,t_1)$ is the indirect utility from choosing y when the period 1 parameter is t_1, given that the value of x will then by chosen optimally given y and t_1. In particular, $y_1 = y^{EX}(F)$ is chosen in period 0 to maximize the expectation $\int g(y,t_1)\, dF(t_1)$.

is SM in (x, y) and has increasing differences in $(x, y; t)$, then f is SM in (x, y, t). You can use this fact without proving it.)

(b) Under the assumptions given above that f is supermodular in (x, y) and has increasing differences in $(x, y; t)$, if $G \geq_{FOSD} F$ then do we necessarily have $y^{EX}(G) \geq y^{EX}(F)$? If $G \geq_{MLR} F$ then do we necessarily have $y^{EX}(G) \geq y^{EX}(F)$? For each of these questions, if your answer is yes, prove your claim. If no, provide a sufficient condition that would generate monotonicity of y with respect to these stochastic orders.

(c) Assume that $y^{EX}(G) \geq y^{EX}(F)$ whenever $G \geq_{FOSD} F$ (that is, regardless of your answer to part (b), assume now that we have imposed whatever conditions are needed to ensure this monotonicity of $y^{EX}(F)$). Fix $\bar{t}' > \bar{t}$, suppose $t_0 = \bar{t}$, and suppose the cdf F puts probability $\alpha \in (0, 1)$ on $t_1 = \bar{t}$ and probability $1 - \alpha$ on $t_1 = \bar{t}'$. Answer the following based on the information provided, and provide a careful argument in support of your answers:

   i. If the realization of t_1 is $\bar{t}'$, what equalities or inequalities can you infer about how $y_0 = y^{LR}(t_0)$, $y_1 = y^{EX}(F)$, and $y_2 = y^{LR}(t_1)$ are related, and about how $x_0 = x^{LR}(t_0)$, $x_1 = x^{SR}(y_1, t_1)$, and $x_2 = x^{LR}(t_1)$ are related?

   ii. If the realization of t_1 is $\bar{t}$, what equalities or inequalities can you infer about how $y_0 = y^{LR}(t_0)$, $y_1 = y^{EX}(F)$, and $y_2 = y^{LR}(t_1)$ are related, and about how $x_0 = x^{LR}(t_0)$, $x_1 = x^{SR}(y_1, t_1)$, and $x_2 = x^{LR}(t_1)$ are related?

(d) Discuss briefly how your results for this problem compare to the conclusions derived in class when we analyzed the Le Chatelier principle. In this setting where the agent is not completely surprised by the change in the parameter t and instead has belief F about its future value, can we still conclude that the long-run change in x is larger than the short-run change in x following a change in the parameter t?

Chapter 6

Modeling and Comparing Information

Contents

References and Assigned Readings

Primary readings:

(1) de Oliveira (2018) (Sections 1–4 recommended)
(2) Mas-Colell, Whinston, and Green (1995), Section 19.B (some basics of modeling information using trees and partitions) (recommended).
(3) Mas-Colell, Whinston, and Green (1995), Section 19.H (imperfect information, comparing information content of deterministic signals, possible detrimental effects of information in strategic contexts) (recommended).

Additional references:

(1) Blackwell (1951, 1953)
(2) Gollier (2001) (Chapter 24)

6.1 Deterministic Signals and Partitions

To introduce concepts gradually, we begin our analysis in this section with the simple case of deterministic signals. Then, in the following sections we develop results for the more general case of stochastic signals. Throughout the chapter, we will restrict attention to a finite probability space. That is, suppose that the state space Ω is finite and P is a probability measure on Ω. It is worth emphasizing that this assumption of finite states is made solely for expositional simplicity. Most of the results presented below extend to probability spaces with an infinite set of states.

6.1.1 Modeling Information Using Deterministic Signals

Definition 6.1. A **(deterministic) signal** $\sigma : \Omega \to S$ maps from the state space Ω to a signal space S.

Note that we refer to the mapping σ as the *signal*, and we refer to s as the *signal realization*.[1] Given a prior P and a signal σ, let P_σ denote the resulting joint probability distribution over Ω and S: The joint probability of state ω and signal realization s is

$$P_\sigma(\omega, s) = \begin{cases} P(\omega) & \text{if } \sigma(\omega) = s \\ 0 & \text{if } \sigma(\omega) \neq s. \end{cases}$$

The unconditional probability of observing signal realization s is the marginal of this joint

[1] Different authors adopt different terminology. For example, some authors refer to σ as the *signal function*, *information structure*, or *experiment*. In these cases, the signal realization s might sometimes be referred to simply as the *signal*.

probability:

$$P_\sigma(s) = \sum_{\omega\in\Omega} P_\sigma(\omega, s) = \sum_{\omega:\sigma(\omega)=s} P(\omega).$$

The posterior probability of state ω given signal realization s is

$$P_\sigma(\omega|s) = \frac{P_\sigma(\omega,s)}{P_\sigma(s)} = \begin{cases} \dfrac{P(\omega)}{\sum_{\omega':\sigma(\omega')=s} P(\omega')} & \text{if } \sigma(\omega)=s \\ 0 & \text{if } \sigma(\omega)\neq s, \end{cases}$$

provided $P_\sigma(s) > 0$. We will discuss posterior beliefs in more detail in later sections.

Definition 6.2. A **partition** of Ω is a collection $\mathcal{E}$ of nonempty disjoint subsets whose union is Ω. That is, $\bigcup_{E\in\mathcal{E}} E = \Omega$ and for any $E, E' \in \mathcal{E}$, either $E = E'$ or $E \cap E' = \emptyset$.

For any subset $S' \subseteq S$, the inverse image of S' under σ is the set

$$\sigma^{-1}(S') \equiv \{\omega \in \Omega : \sigma(\omega) \in S'\}.$$

Note that the inverse image is a well-defined (but possibly empty) set even if σ is not a bijection. That is, even if a function does not admit an inverse function, we can still define the inverse image of sets under that function.

The partition generated by the signal σ is the collection

$$\mathcal{E} = \{\sigma^{-1}(\{s\}) \subseteq \Omega : s \in S^*\},$$

where S^* is the range of σ. Thus if $\omega \in E \in \mathcal{E}$, then

$$E = \{\omega' \in \Omega : \sigma(\omega') = \sigma(\omega)\}$$

and hence

$$P_\sigma(\omega'|\sigma(\omega)) = P(\omega'|E).$$

Definition 6.3. Suppose $\mathcal{E}_1$ and $\mathcal{E}_2$ are partitions of Ω. Then $\mathcal{E}_1$ is **finer** than $\mathcal{E}_2$ (or, equivalently, $\mathcal{E}_2$ is **coarser** than $\mathcal{E}_1$) if every element of $\mathcal{E}_1$ is a subset of some element of $\mathcal{E}_2$, that is, for every $E \in \mathcal{E}_1$, there exists $E' \in \mathcal{E}_2$ such that $E \subseteq E'$.

Example 6.4 (Deterministic signals). Suppose $\Omega = \{1, 2, 3, 4, 5\}$ and $S = \{a, b, c\}$, and consider the following two signals:

$$\sigma_1(\omega) = \begin{cases} a & \text{if } \omega = 1 \\ a & \text{if } \omega = 2 \\ b & \text{if } \omega = 3 \\ c & \text{if } \omega = 4 \\ c & \text{if } \omega = 5 \end{cases} \qquad \sigma_2(\omega) = \begin{cases} c & \text{if } \omega = 1 \\ c & \text{if } \omega = 2 \\ c & \text{if } \omega = 3 \\ a & \text{if } \omega = 4 \\ a & \text{if } \omega = 5 \end{cases}$$

We can also depict this two signals using tables:

$\sigma_1(\omega)$

	a	b	c
1	x		
2	x		
3		x	
4			x
5			x

$\sigma_2(\omega)$

	a	b	c
1			x
2			x
3			x
4	x		
5	x		

These signals generate the partitions

$$\mathcal{E}_1 = \{\{1,2\},\{3\},\{4,5\}\} \qquad \text{and} \qquad \mathcal{E}_2 = \{\{1,2,3\},\{4,5\}\}.$$

Notice that $\mathcal{E}_1$ is finer than $\mathcal{E}_2$. Notice also that the information content in signal 1 is greater than that of signal 2 in the following intuitive sense: If an individual observes signal 1, then they can replicate signal 2 with perfect accuracy; that is, there is a mapping $\gamma : S \to S$ such that $\sigma_2 = \gamma \circ \sigma_1$. It is also worth observing that not every partition is comparable in this way. For instance, if there is a third signal σ_3 that generates that partition $\mathcal{E}_3 = \{\{1,2\},\{3,4,5\}\}$, then $\mathcal{E}_1$ is finer than $\mathcal{E}_3$, but neither $\mathcal{E}_2$ nor $\mathcal{E}_3$ is finer than the other. Thus, comparing partitions based on which is finer induces a partial order on the set of all partitions.

We will show momentarily that the connection between one signal generating a finer partition than another and the ability to replicate the second signal after observing the first is not special to the previous example, but an equivalence that holds in general. First, we introduce feasible action plans and the ex-ante expected utility that can be generated by signals.

6.1.2 State-Contingent Action Plans

Definition 6.5. Fix a set of actions A. A state-contingent action plan $\lambda : \Omega \to A$ is **feasible** given the signal $\sigma : \Omega \to S$ if there exists a function $\alpha : S \to A$ such that $\lambda = \alpha \circ \sigma$, that is, $\lambda(\omega) = \alpha(\sigma(\omega))$ for all ω.

Equivalently, λ is feasible if $\sigma(\omega) = \sigma(\omega')$ implies $\lambda(\omega) = \lambda(\omega')$ (Exercise 6.1). Let Λ_σ

denote the set of all feasible state-contingent action plans given the signal σ, so

$$\Lambda_\sigma = \{\lambda : \Omega \to A : \lambda = \alpha \circ \sigma \text{ for some } \alpha : S \to A\}.$$

While our notation emphasizes that Λ_σ depends on the signal σ, keep in mind that this set also depends on the set of available actions A.

Definition 6.6. Consider any Bayesian expected-utility maximizer with a set of actions A, state-dependent utility function $u : A \times \Omega \to \mathbb{R}$, and a prior P on Ω. The individual's **ex-ante expected utility** from the signal σ is

$$\max_{\lambda \in \Lambda_\sigma} \sum_{\omega \in \Omega} u(\lambda(\omega), \omega)\, P(\omega) = \max_{\alpha : S \to A} \sum_{\omega \in \Omega} u(\alpha(\sigma(\omega)), \omega)\, P(\omega).$$

6.1.3 Comparing Information

When is one deterministic signal more informative than another? There are several natural ways to think about this comparison, such whether one of them: generates a finer partition, can be used to replicate the other, induces a larger set of feasible state-contingent action plans, or generates higher ex-ante expected utility. It turns out that all of these methods of comparison are equivalent, as the following theorem demonstrates.

Theorem 6.7. Suppose $\sigma_1 : \Omega \to S_1$ and $\sigma_2 : \Omega \to S_2$ are two signals (with possibly different signal spaces). The following are equivalent:

(1) The partition $\mathcal{E}_1$ generated by σ_1 is finer than the partition $\mathcal{E}_2$ generated by σ_2.
(2) There exists a function $\gamma : S_1 \to S_2$ such that $\sigma_2 = \gamma \circ \sigma_1$. That is, $\sigma_2(\omega) = \gamma(\sigma_1(\omega))$ for all ω.
(3) For any set of actions A, the set of state-contingent action plans that are feasible under σ_1 contains those that are feasible under σ_2. That is, $\Lambda_{\sigma_2} \subseteq \Lambda_{\sigma_1}$.
(4) Every Bayesian expected-utility maximizer prefers σ_1 to σ_2 for every possible decision problem. That is, σ_1 gives weakly higher ex-ante utility than σ_2 for every A, u, and P.

Proof. (1) $\Leftrightarrow$ (2): Exercise 6.1.

(2) $\Rightarrow$ (3): Suppose $\lambda \in \Lambda_{\sigma_2}$. Then there exists a function $\alpha : S_2 \to A$ such that $\lambda = \alpha \circ \sigma_2$. By assumption, there exists a function $\gamma : S_1 \to S_2$ such that $\sigma_2 = \gamma \circ \sigma_1$. Let $\alpha^* = \alpha \circ \gamma : S_1 \to A$. Then

$$\alpha^* \circ \sigma_1 = (\alpha \circ \gamma) \circ \sigma_1 = \alpha \circ (\gamma \circ \sigma_1) = \alpha \circ \sigma_2 = \lambda,$$

so $\lambda \in \Lambda_{\sigma_1}$. Thus $\Lambda_{\sigma_2} \subseteq \Lambda_{\sigma_1}$.

(3) $\Rightarrow$ (4): Fix any set of actions A, state-dependent utility function $u : A \times \Omega \to \mathbb{R}$, and

prior P on Ω. By assumption, $\Lambda_{\sigma_2} \subseteq \Lambda_{\sigma_1}$, and therefore

$$\max_{\lambda \in \Lambda_{\sigma_2}} \sum_{\omega \in \Omega} u(\lambda(\omega), \omega)\, P(\omega) \leq \max_{\lambda \in \Lambda_{\sigma_1}} \sum_{\omega \in \Omega} u(\lambda(\omega), \omega)\, P(\omega).$$

(4) $\Rightarrow$ (2): Consider a utility function that is maximized by announcing the signal realization under σ_2. If σ_1 performs better in this decision problem, then it must be possible to replicate the signal realization given by σ_2. Formally, let $A = S_2$, define u by

$$u(s, \omega) = \begin{cases} 1 & \text{if } s = \sigma_2(\omega) \\ 0 & \text{if } s \neq \sigma_2(\omega), \end{cases}$$

and let P be the uniform distribution on Ω. By construction,

$$\sum_{\omega \in \Omega} u(\lambda(\omega), \omega)\, P(\omega) \leq 1,$$

with equality if and only if $\lambda = \sigma_2$. Note that $\sigma_2 \in \Lambda_{\sigma_2}$ and therefore

$$\max_{\lambda \in \Lambda_{\sigma_2}} \sum_{\omega \in \Omega} u(\lambda(\omega), \omega)\, P(\omega) = 1.$$

Since the individual gets (weakly) higher utility from σ_1 than σ_2, it must be that $\sigma_2 \in \Lambda_{\sigma_1}$. That is, $\sigma_2 = \alpha \circ \sigma_1$ for some $\alpha : S_1 \to S_2$. $\square$

6.2 Stochastic Signals

We now extend the analysis of the previous section to the general case where signal realizations are potentially random. For all of our results (and most of our examples), we will continue to assume that the state space Ω is finite and that the signal space S is finite. As noted above, the reader should keep in mind that these restrictions are made for expositional simplicity and significant generalizations exist in the literature.

6.2.1 Modeling Information Using Stochastic Signals

For any finite set X, let

$$\triangle(X) = \left\{ p \in \mathbb{R}_+^X : \sum_{x \in X} p(x) = 1 \right\}$$

denote the set of all probability distributions on X. In particular, a stochastic signal will be a mapping from states into $\triangle(S)$.

Definition 6.8. A **(stochastic) signal** $\sigma : \Omega \to \triangle(S)$ maps from the state space Ω to the set of probability distributions over a signal space S.

We will write $\sigma(s|\omega)$ to denote the probability of observing s when the state is ω. Note that a deterministic signal is the special case of a stochastic signal where, for each $\omega \in \Omega$, we have $\sigma(s|\omega) = 1$ for some signal realization $s \in S$ (the particular s may of course depend on the state ω). Therefore, we will often refer to a stochastic signal simply as a **signal** in what follows, recognizing that this encompasses deterministic signals.

Note that the information structure associated with a stochastic signal in general cannot be modeled using a partition on Ω. However, it can be thought of as a partition on the enlarged state space $\Omega^* = \Omega \times S$. That is,

$$\mathcal{E} = \{\Omega \times \{s\} : s \in S\}.$$

Unfortunately, this is not particularly useful for applications (when only Ω is payoff relevant) or for comparing information. For example, two individuals may have different stochastic signals σ_1 and σ_2 with corresponding signal spaces S_1 and S_2. Then the enlarged state space is $\Omega^* = \Omega \times S_1 \times S_2$, and the partitions generated by σ_1 and σ_2 are

$$\mathcal{E}_1 = \{\Omega \times \{s_1\} \times S_2 : s_1 \in S_1\} \quad \text{and} \quad \mathcal{E}_2 = \{\Omega \times S_1 \times \{s_2\} : s_2 \in S_2\}.$$

In this case, neither partition is finer than the other. However, as we will explore in detail, σ_1 may still be more informative about Ω than σ_2. Intuitively, if Ω is the only payoff-relevant part of this state space, then individual 1 does not need to know individual 2's signal realization in order to have more information than 2 for any relevant decision problem.

EXAMPLE 6.9 (STOCHASTIC EXTENSION OF DETERMINISTIC EXAMPLE). Consider a simple stochastic extension of the deterministic signals in Example 6.4: Suppose $\Omega = \{1, 2, 3, 4, 5\}$ and $S = \{a, b, c, d\}$. Suppose the signals σ_1 and σ_2 announce the same signal realizations as their counterparts from the previous example with probability 0.5, and they announce signal realization d with probability 0.5. These signals can be conveniently represented using the following tables:

$\sigma_1(s_1|\omega)$

	a	b	c	d
1	0.5			0.5
2	0.5			0.5
3		0.5		0.5
4			0.5	0.5
5			0.5	0.5

$\sigma_2(s_2|\omega)$

	a	b	c	d
1			0.5	0.5
2			0.5	0.5
3			0.5	0.5
4	0.5			0.5
5	0.5			0.5

Notice that for either signal, after observing the signal realization d an individual has no information about the state beyond their prior belief P. Also, notice that unlike in the deterministic example, observing the realization of signal 1 does not enable an individual to announce the realization of signal 2 with perfect accuracy. For instance, suppose the state is $\omega = 1$ and the first signal has realization a. This does not fully pin down the realization of the second signal, since it could be either c or d, each with equal probability. However, signal 1 can be used to replicate signal 2 in a distributional sense. Specifically, suppose that after observing the realization of signal σ_1, an individual makes the following announcements:

- announce c after observing either a or b,
- announce a after observing c, and
- announce d after observing d.

These signal announcements depend only on the realization of the first signal and result in the same distribution of signal realizations in every state ω as under signal σ_2. We will see shortly that this type of comparison of stochastic signals is the appropriate one for ranking ex-ante expected utilities.

To formalize the notion of using one stochastic signal to replicate another, we need to first define the composition of stochastic operators. Suppose $\alpha : X \to \triangle(Y)$ and $\beta : Y \to \triangle(Z)$ are two stochastic operators. Define the **composition** $\beta \circ \alpha : X \to \triangle(Z)$ of these operators by

$$(\beta \circ \alpha)(z|x) = \sum_{y \in Y} \beta(z|y)\alpha(y|x).$$

We now define a garbling of a signal, which can be interpreted intuitively as adding more noise to the signal.

Definition 6.10. We say that a stochastic signal $\sigma_2 : \Omega \to \triangle(S_2)$ is a **garbling** of $\sigma_1 : \Omega \to \triangle(S_1)$ (or alternatively that σ_1 is **Blackwell sufficient** for σ_2) if there exists a function $\gamma : S_1 \to \triangle(S_2)$ (the garbling) such that $\sigma_2 = \gamma \circ \sigma_1$. That is, for all $\omega \in \Omega$ and $s_2 \in S_2$,

$$\sigma_2(s_2|\omega) = \sum_{s_1 \in S_1} \gamma(s_2|s_1)\sigma_1(s_1|\omega).$$

In Example 6.9, the garbling used to obtain σ_2 from σ_1 turned out to be deterministic: Specifically, let

$$\gamma(c|a) = \gamma(c|b) = \gamma(a|c) = \gamma(d|d) = 1,$$

and let γ take the value zero for all other combinations of s_1 and s_2. However, in other cases, such as the following example, the garbling is nontrivially stochastic.

EXAMPLE 6.11 (BINARY STATE AND SIGNAL SPACES). Suppose $\Omega = \{G, B\}$ and $S = \{g, b\}$, and consider the following signals:

$\sigma_1(s_1|\omega)$

	g	b
G	0.8	0.2
B	0.2	0.8

$\sigma_2(s_2|\omega)$

	g	b
G	0.6	0.4
B	0.4	0.6

We will show that signal 2 is a garbling of signal 1. Intuitively, we can obtain the second signal from the first by announcing the same signal realization as under σ_1 (g if g and b if b) with probability 2/3 and announcing the opposite signal realization (b if g and g if b) with probability 1/3. This garbling is summarized in the following table, where the rows correspond to the signal realization s_1 and the columns correspond to the signal realization s_2:

$\gamma(s_2|s_1)$

		s_2: g	s_2: b
s_1	g	2/3	1/3
	b	1/3	2/3

To verify that we do in fact have $\sigma_2 = \gamma \circ \sigma_1$, notice that

$$\begin{aligned}(\gamma \circ \sigma_1)(g|G) &= \gamma(g|g)\sigma_1(g|G) + \gamma(g|b)\sigma_1(b|G) \\ &= (2/3)0.8 + (1/3)0.2 = 0.6 = \sigma_2(g|G),\end{aligned}$$

and similarly for the other combinations of ω and s_2 (which you should check for yourself). Thus, σ_2 is a garbling of σ_1. While it may have seemed somewhat mysterious how we came up with this garbling γ, we will explore both the geometric intuition and algebraic calculations behind such derivations in more detail as the chapter progresses and in the exercises.

The tables used to describe the signals and the garbling in the previous example are evocative of matrices in linear algebra, and there is indeed a formal connection. If the sets Ω, S_1, and S_2 are finite (as we assume for all of the formal results of this chapter), then signals and the garbling can be representing as matrices, and the garbling equation can be represented using matrix multiplication. Using the signals from Example 6.11 to illustrate, with slight abuse of notation, we can treat σ_1, σ_2, and γ as the following matrices:

$$\sigma_1 = \begin{bmatrix} 0.8 & 0.2 \\ 0.2 & 0.8 \end{bmatrix}, \qquad \sigma_2 = \begin{bmatrix} 0.6 & 0.4 \\ 0.4 & 0.6 \end{bmatrix}, \qquad \gamma = \begin{bmatrix} 2/3 & 1/3 \\ 1/3 & 2/3 \end{bmatrix}.$$

Moreover, $\gamma \circ \sigma_1$ can be represented as the matrix multiplication $\sigma_1\gamma$:

$$\gamma \circ \sigma_1 = \begin{bmatrix} 0.8 & 0.2 \\ 0.2 & 0.8 \end{bmatrix} \begin{bmatrix} 2/3 & 1/3 \\ 1/3 & 2/3 \end{bmatrix} = \begin{bmatrix} 0.6 & 0.4 \\ 0.4 & 0.6 \end{bmatrix} = \sigma_2.$$

In other words, the signal resulting from the garbling operation can be described using the standard composition of stochastic transition matrices (Markov kernels).

6.2.2 State-Contingent Distributions Over Actions

The definition of state-contingent actions is roughly the same as before, except that we now allow both the signal σ and the action α to be stochastic.

Definition 6.12. Fix a set of actions A. A state-contingent distribution over actions $\lambda : \Omega \to \triangle(A)$ is **feasible** given the signal $\sigma : \Omega \to \triangle(S)$ if there exists a distribution of actions conditional on the signal realization $\alpha : S \to \triangle(A)$ such that $\lambda = \alpha \circ \sigma$. That is,

$$\lambda(a|\omega) = \sum_{s \in S} \alpha(a|s)\sigma(s|\omega).$$

The set of all feasible state-contingent distributions over actions is then

$$\Lambda_\sigma = \{\lambda : \Omega \to \triangle(A) : \lambda = \alpha \circ \sigma \text{ for some } \alpha : S \to \triangle(A)\}.$$

While our notation emphasizes that Λ_σ depends on the signal σ, keep in mind that this set also depends on the set of available actions A.

Definition 6.13. Consider a Bayesian expected-utility maximizer with a set of actions A, state-dependent Bernoulli utility function $u : A \times \Omega \to \mathbb{R}$, and a prior P on Ω. The individual's **ex-ante expected utility** from the signal σ is

$$\max_{\lambda \in \Lambda_\sigma} \sum_{\omega \in \Omega} \left(\sum_{a \in A} u(a,\omega)\lambda(a|\omega) \right) P(\omega).$$

Although it will generally be more convenient for our purposes to express ex-ante expected utility as a function of state-contingent action plans λ, it is of course also possible to express utility as a function of signal-contingent action plans α using the following equivalent formulation:

$$\max_{\alpha : S \to \triangle(A)} \sum_{\omega \in \Omega} \left(\sum_{a \in A} \sum_{s \in S} u(a,\omega)\alpha(a|s)\sigma(s|\omega) \right) P(\omega).$$

Note that because expected utility is linear in α, it is easy to see that there will always be a maximizer involving deterministic α (this may still result in random λ due to the stochastic signal). In this sense, we could restrict attention to deterministic $\alpha : S \to A$. However, taking $\alpha : S \to \triangle(A)$ will be technically convenient for the analysis because it ensures that the set Λ_σ is convex (this will be used in the proof of Blackwell's theorem below).

After presenting Blackwell's theorem in the next subsection, we will revisit Example 6.11 in Section 6.3 and compare both the sets of feasible state-contingent distributions of actions and the ex-ante expected utilities for the two signals for given A, u, and P.

6.2.3 Comparing Information and Blackwell's Theorem

We now present our main result for the comparison of stochastic signals.

Theorem 6.14 (Blackwell (1951, 1953)). Suppose $\sigma_1 : \Omega \to \triangle(S_1)$ and $\sigma_2 : \Omega \to \triangle(S_2)$

are two stochastic signals. The following are equivalent:

(1) σ_2 is a garbling of σ_1.
(2) For any set of actions A, the set of state-contingent distributions over actions that are feasible under σ_1 contains those that are feasible under σ_2. That is, $\Lambda_{\sigma_2} \subseteq \Lambda_{\sigma_1}$.
(3) Every Bayesian expected-utility maximizer prefers σ_1 to σ_2 for any possible decision problem. That is, σ_1 gives weakly higher ex-ante utility than σ_2 for every A, u, and P.

While the classical proofs of Blackwell's theorem are quite involved, we provide a very simple proof that was developed by de Oliveira (2018). Graphical intuition for some parts of this proof, in particular for (2) $\Leftrightarrow$ (3), will be provided using examples in Section 6.3.

Proof. (1) $\Rightarrow$ (2): Suppose $\lambda \in \Lambda_{\sigma_2}$. Then there exists $\alpha : S_2 \to \triangle(A)$ such that $\lambda = \alpha \circ \sigma_2$. By assumption, there exists a garbling $\gamma : S_1 \to \triangle(S_2)$ such that $\sigma_2 = \gamma \circ \sigma_1$. Let $\alpha^* = \alpha \circ \gamma : S_1 \to \triangle(A)$. Then

$$\alpha^* \circ \sigma_1 = (\alpha \circ \gamma) \circ \sigma_1 = \alpha \circ (\gamma \circ \sigma_1) = \alpha \circ \sigma_2 = \lambda.$$

The key step in this equation is using the associative property of our composition operation to obtain $(\alpha \circ \gamma) \circ \sigma_1 = \alpha \circ (\gamma \circ \sigma_1)$. To verify this equality, note that

$$\begin{aligned}
((\alpha \circ \gamma) \circ \sigma_1)(a|\omega) &= \sum_{s_1 \in S_1} \underbrace{\left(\sum_{s_2 \in S_2} \alpha(a|s_2)\gamma(s_2|s_1) \right)}_{(\alpha\circ\gamma)(a|s_1)} \sigma_1(s_1|\omega) \\
&= \sum_{s_2 \in S_2} \alpha(a|s_2) \underbrace{\left(\sum_{s_1 \in S_1} \gamma(s_2|s_1)\sigma_1(s_1|\omega) \right)}_{(\gamma\circ\sigma_1)(s_2|\omega)} \\
&= (\alpha \circ (\gamma \circ \sigma_1))(a|\omega).
\end{aligned}$$

Thus, $\lambda \in \Lambda_{\sigma_1}$. Since $\lambda \in \Lambda_{\sigma_2}$ was arbitrary, we have shown that $\Lambda_{\sigma_2} \subseteq \Lambda_{\sigma_1}$.

(2) $\Rightarrow$ (1): Let $A = S_2$. Then, $\sigma_2 \in \Lambda_{\sigma_2}$. Since $\Lambda_{\sigma_2} \subseteq \Lambda_{\sigma_1}$, it must be that $\sigma_2 \in \Lambda_{\sigma_1}$. That is, $\sigma_2 = \alpha \circ \sigma_1$ for some $\alpha : S_1 \to \triangle(S_2)$.

(2) $\Rightarrow$ (3): Fix any set of actions A, state-dependent utility function $u : A \times \Omega \to \mathbb{R}$, and prior P on Ω. By assumption, $\Lambda_{\sigma_2} \subseteq \Lambda_{\sigma_1}$, and therefore

$$\max_{\lambda \in \Lambda_{\sigma_2}} \sum_{\omega \in \Omega} \left(\sum_{a \in A} u(a,\omega)\lambda(a|\omega) \right) P(\omega) \leq \max_{\lambda \in \Lambda_{\sigma_1}} \sum_{\omega \in \Omega} \left(\sum_{a \in A} u(a,\omega)\lambda(a|\omega) \right) P(\omega).$$

(3) $\Rightarrow$ (2): We will prove this by contrapositive. That is, we will show that if (2) fails then (3) fails. Thus, suppose that (2) is not true, so there exists a set of actions A such

that $\Lambda_{\sigma_2} \not\subseteq \Lambda_{\sigma_1}$. Fix some $\lambda^* \in \Lambda_{\sigma_2}$ such that $\lambda^* \notin \Lambda_{\sigma_1}$. Note that we can think of λ^* as an element of the Euclidean space $\mathbb{R}^{A\times\Omega}$ (that is, the real-valued vectors that have coordinates indexed by (ω, a)), and similarly Λ_{σ_1} is a subset of $\mathbb{R}^{A\times\Omega}$. Note that Λ_{σ_1} is convex (can you show why?) and compact (why?). Therefore, the separating hyperplane theorem implies that there exists a vector $u \in \mathbb{R}^{A\times\Omega}$ such that

$$\sum_{\substack{\omega\in\Omega \\ a\in A}} u(a,\omega)\lambda(a|\omega) < \sum_{\substack{\omega\in\Omega \\ a\in A}} u(a,\omega)\lambda^*(a|\omega) \qquad \forall \lambda \in \Lambda_{\sigma_1}.$$

Let P be the uniform distribution on Ω, so that $P(\omega) = 1/|\Omega|$ for all ω. Then, we have

$$\begin{aligned}\max_{\lambda\in\Lambda_{\sigma_1}} \sum_{\omega\in\Omega}\left(\sum_{a\in A} u(a,\omega)\lambda(a|\omega)\right)P(\omega) &< \sum_{\omega\in\Omega}\left(\sum_{a\in A} u(a,\omega)\lambda^*(a|\omega)\right)P(\omega) \\ &\leq \max_{\lambda\in\Lambda_{\sigma_2}} \sum_{\omega\in\Omega}\left(\sum_{a\in A} u(a,\omega)\lambda(a|\omega)\right)P(\omega),\end{aligned}$$

and hence (3) fails. This completes the proof. □

Definition 6.15. If any of the conditions in Theorem 6.14 hold, we say that σ_1 is **Blackwell more informative** (or simply more informative) than σ_2.

6.3 Additional Examples

The following examples illustrate Blackwell's theorem. Note in these examples that the prior P is not needed when describing garblings. This is because the Blackwell comparison of the informativeness of information structures (signals) does not depend on the prior.

6.3.1 Binary State and Signal Spaces

We now revisit our previous example involving binary state and signal spaces to illustrate the connection between the first two conditions in Theorem 6.14.

EXAMPLE 6.16 (DEPICTING STATE-CONTINGENT ACTIONS GRAPHICALLY). Continuing Example 6.11, suppose there are two actions $A = \{x, y\}$. For example, x could be high investment in a project and y could be low investment. In this simple binary state, signal, and action space environment, we can depict the sets of feasible state-contingent distributions over actions Λ_{σ_1} and Λ_{σ_2} as subset of $\mathbb{R}^2$. That is, since for any λ we have

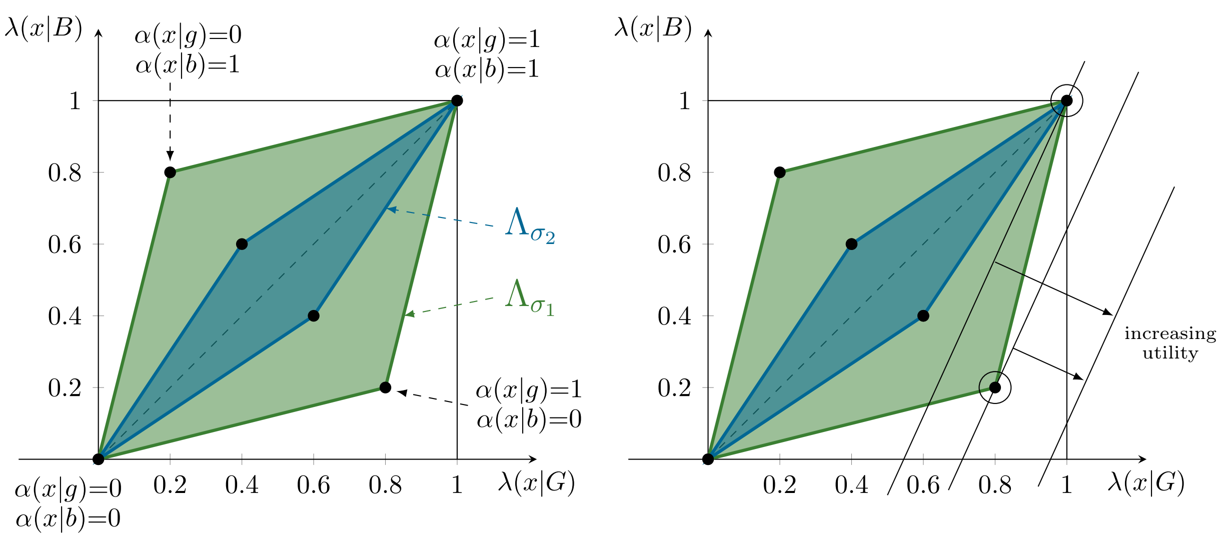
$$\lambda(y|G) = 1 - \lambda(x|G) \qquad \text{and} \qquad \lambda(y|B) = 1 - \lambda(x|B),$$

a state-contingent distribution is completely pinned down by the values $\lambda(x|G)$ and $\lambda(x|B)$.

The following tables illustrate these values for each of the four *deterministic* signal-contingent action plans α given the signals σ_1 and σ_2:

	Signal σ_1		Signal σ_2	
	$\lambda(x\|G)$	$\lambda(x\|B)$	$\lambda(x\|G)$	$\lambda(x\|B)$
$\alpha(x\|g) = 0,\ \alpha(x\|b) = 0$	0	0	0	0
$\alpha(x\|g) = 1,\ \alpha(x\|b) = 0$	0.8	0.2	0.6	0.4
$\alpha(x\|g) = 0,\ \alpha(x\|b) = 1$	0.2	0.8	0.4	0.6
$\alpha(x\|g) = 1,\ \alpha(x\|b) = 1$	1	1	1	1

For example, for signal σ_1, the λ resulting from choosing x after signal realization g (i.e., $\alpha(x|g) = 1$) and choosing y after signal realization b (i.e., $\alpha(x|b) = 0$) can be represented using the vector $(0.8, 0.2)$. What about other, non-deterministic signal-contingent distributions over actions? Any $\alpha : S \to \triangle(A)$ must be a convex combination of the four deterministic functions α listed above, and therefore the resulting state-contingent distribution of actions λ is simply the convex combination (using the same weights) of the entries in the table. In other words, the set Λ_{σ_i} is just the convex combination of the four extreme points listed in the table. This allows us to easily calculate and depict the sets Λ_{σ_1} and Λ_{σ_2} graphically, as illustrated in Figure 6.1a. It is immediate that $\Lambda_{\sigma_2} \subseteq \Lambda_{\sigma_1}$.



(a) State-contingent distributions over actions

(b) Ex-ante expected utility

Figure 6.1: Illustration of comparable signals in Examples 6.16 and 6.17.

To better understand the third condition in Theorem 6.14, it is helpful to be able to depict ex-ante expected utility graphically. We now continue the previous example and introduce utility functions.

Example 6.17 (Depicting expected-utility functions graphically). Continuing Example 6.16, suppose that $P(G) = P(B) = 0.5$. Note that this assumption is made for simplicity only and is not crucial for any of the analysis to follow (check for yourself what would be different and convince yourself that there are no substantive changes). Then, for any state-contingent distribution over actions λ, we have

$$\begin{aligned}
2\sum_{\omega\in\Omega}\sum_{a\in A} u(a,\omega)\lambda(a|\omega)P(\omega) &= \sum_{\omega\in\Omega}\sum_{a\in A} u(a,\omega)\lambda(a|\omega) \\
&= u(x,G)\lambda(x|G) + u(y,G)(1-\lambda(x|G)) + u(x,B)\lambda(x|B) + u(y,B)(1-\lambda(x|B)) \\
&= \big[u(x,G)-u(y,G)\big]\lambda(x|G) + \big[u(x,B)-u(y,B)\big]\lambda(x|B) + u(y,G) + u(y,B).
\end{aligned}$$

The last two terms in this equation do not depend on λ, so the ranking of any two state-contingent distributions λ and λ' is determined entirely by the utility differences between action x and y in states G and B. For example, suppose x corresponds to high investment in a project and y to low. Then, if high investment is better than low only in the good state G, then we would have

$$u(x,G)-u(y,G) > 0 \qquad \text{and} \qquad u(x,B)-u(y,B) < 0.$$

Moreover, the vector given by these two utility differences is precisely the gradient of the ex-ante expected-utility function, and it is therefore normal to the (linear) indifference curves in the space of action plans λ. Indifference curves for one such utility function, and the resulting optimal state-contingent distributions over actions from Λ_{σ_1} and Λ_{σ_2} are illustrated in Figure 6.1b.

In the previous example, signal 1 was more informative than signal 2. We next consider an example where signals are not Blackwell comparable.

EXAMPLE 6.18 (INCOMPARABLE SIGNALS). Suppose again that $\Omega = \{G, B\}$ and $S = \{g, b\}$, but consider the following signals:

$\sigma_1(s_1|\omega)$

	g	b
G	1	0
B	0.3	0.7

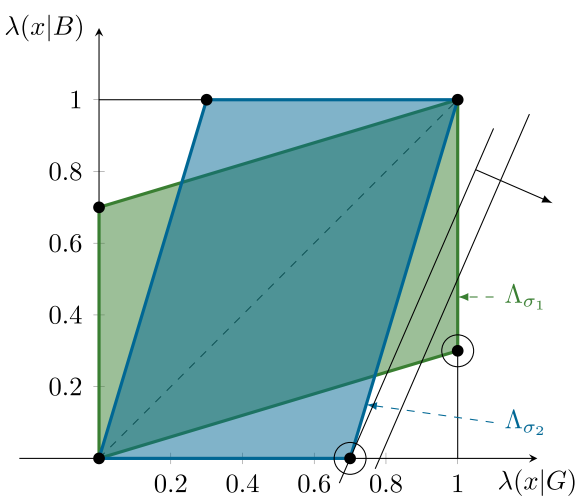
$\sigma_2(s_2|\omega)$

	g	b
G	0.7	0.3
B	0	1

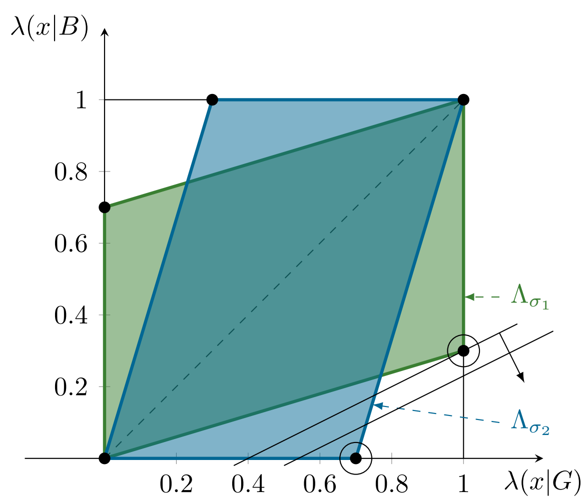
We will show that neither of these signals is more informative than the other in the sense of Blackwell. In light of Theorem 6.14, there are a number of ways to demonstrate this. For instance, we could try to show that neither is a garbling of the other. However, building on the graphical intuition developed in the previous examples, we instead show that there exists a set of actions A such that the sets of state-contingent distributions over actions for these signals are not nested; that is, neither is a subset of the other. Again let $A = \{x, y\}$ and recall that the sets Λ_{σ_1} and Λ_{σ_2} can be represented as subsets of $\mathbb{R}^2$. In particular, each set is the convex hull of the vectors given by the rows in its table below:

	Signal σ_1		Signal σ_2	
	$\lambda(x\|G)$	$\lambda(x\|B)$	$\lambda(x\|G)$	$\lambda(x\|B)$
$\alpha(x\|g) = 0,\ \alpha(x\|b) = 0$	0	0	0	0
$\alpha(x\|g) = 1,\ \alpha(x\|b) = 0$	1	0.3	0.7	0
$\alpha(x\|g) = 0,\ \alpha(x\|b) = 1$	0	0.7	0.3	1
$\alpha(x\|g) = 1,\ \alpha(x\|b) = 1$	1	1	1	1

Figure 6.2 depicts these sets Λ_{σ_1} and Λ_{σ_2} graphically. It is easy to see that neither is a subset of the other, and hence condition 2 of Theorem 6.14 fails. Thus, neither signal is Blackwell more informative than the other. Continuing our analysis, we can also demonstrate that there is a decision problem where σ_1 is preferred and another decision problem where σ_2 is preferred. In other words, we will illustrate that condition 3 in Theorem 6.14 fails, which provides some intuition for why $3 \Rightarrow 2$ in this theorem (by showing that whenever condition 2 fails then condition 3 also fails). Take the vector $(1, 0.3)$, which is contained in Λ_{σ_1} but not in Λ_{σ_2}. By the separating hyperplane theorem, there exists a hyperplane that separates this point from the set Λ_{σ_2}. Since hyperplanes correspond to the indifference curves of ex-ante expected-utility preferences (as we observed in Example 6.17), this means that there is a utility function and prior such that signal 1 is preferred to signal 2, as illustrated in Figure 6.2a. Similarly, we can find another decision problem (utility function) where signal 2 is preferred to signal 1, as illustrated in Figure 6.2b.



(a) Signal 1 gives higher utility



(b) Signal 2 gives higher utility

Figure 6.2: Illustration of incomparable signals in Example 6.18.

The following example generalizes the signals introduced in Example 6.11 and describes the garbling used to obtain the second signal from the first. Note that our approach to graphing the set of feasible state-contingent action plans in the previous examples can also be used to graph the set of all garblings of a given signal in the case of binary state and signal spaces. Give this a try in the following example.

EXAMPLE 6.19. Suppose $\Omega = \{G, B\}$ and $S = \{g, b\}$, and consider the signals $\sigma_i : \Omega \to \triangle(S)$ for $i = 1, 2$ defined by

$\sigma_1(s_1|\omega)$

	g	b
G	ϕ_1	$1-\phi_1$
B	$1-\phi_1$	ϕ_1

$\sigma_2(s_2|\omega)$

	g	b
G	ϕ_2	$1-\phi_2$
B	$1-\phi_2$	ϕ_2

If $1/2 < \phi_2 < \phi_1$, then 1 can replicate 2's signal distribution conditional on ω:

- Let $\alpha = \frac{\phi_2+\phi_1-1}{2\phi_1-1} \in (1/2, 1)$. For example, if $\phi_1 = 1$, then $\alpha = \phi_2$.
- When 1 sees signal realization g, announce g with probability $\gamma(g|g) = \alpha$ and b with probability $\gamma(b|g) = 1 - \alpha$.
- When 1 sees signal realization b, announce b with probability $\gamma(b|b) = \alpha$ and g with probability $\gamma(g|b) = 1 - \alpha$.

Note that 1's announcement has the same distribution as 2's signal conditional on both states, G and B:

$$\gamma(g|g)\sigma_1(g|G) + \gamma(g|b)\sigma_1(b|G) = \alpha\phi_1 + (1-\alpha)(1-\phi_1) = \phi_2 = \sigma_2(g|G)$$
$$\gamma(g|g)\sigma_1(g|B) + \gamma(g|b)\sigma_1(b|B) = \alpha(1-\phi_1) + (1-\alpha)\phi_1 = 1-\phi_2 = \sigma_2(g|B).$$

Thus, σ_2 is a garbling of σ_1. As we showed early in the chapter, we can also describe these signals and the garbling using matrices:

$$\sigma_1 = \begin{bmatrix} \phi_1 & 1-\phi_1 \\ 1-\phi_1 & \phi_1 \end{bmatrix}, \qquad \sigma_2 = \begin{bmatrix} \phi_2 & 1-\phi_2 \\ 1-\phi_2 & \phi_2 \end{bmatrix}, \qquad \gamma = \begin{bmatrix} \alpha & 1-\alpha \\ 1-\alpha & \alpha \end{bmatrix}.$$

Recall that $\gamma \circ \sigma_1$ can be represented as the matrix multiplication $\sigma_1\gamma$, and hence:

$$\gamma \circ \sigma_1 = \begin{bmatrix} \phi_1 & 1-\phi_1 \\ 1-\phi_1 & \phi_1 \end{bmatrix} \begin{bmatrix} \alpha & 1-\alpha \\ 1-\alpha & \alpha \end{bmatrix} = \begin{bmatrix} \phi_2 & 1-\phi_2 \\ 1-\phi_2 & \phi_2 \end{bmatrix} = \sigma_2.$$

6.3.2 Location Experiments

The next two examples involve infinite signal spaces S. Although we only stated the definition of a garbling and proved the equivalence of different conditions corresponding to better informativeness in the case of finite signal spaces, the definitions and results can be extended to these more general environments.

EXAMPLE 6.20 (LOCATION EXPERIMENT—NORMAL ERRORS). Assume:

- $\Omega \subseteq \mathbb{R}$ (no need to specify prior)
- $S_1 = S_2 = \mathbb{R}$
- 1 observes $s_1 = \omega + \varepsilon_1$ for $\varepsilon_1 \sim N(0, \eta_1^2)$

- 2 observes $s_2 = \omega + \varepsilon_2$ for $\varepsilon_2 \sim N(0, \eta_2^2)$
- If $\eta_1 \leq \eta_2$, then 1 has better information: Conditional on ω, s_2 is distributed the same as $s_1 + \varepsilon$ for $\varepsilon \sim N(0, \eta_2^2 - \eta_1^2)$, since $s_1 + \varepsilon = \omega + \varepsilon_1 + \varepsilon$ and $\varepsilon_1 + \varepsilon \sim N(0, \eta_2^2)$.

Example 6.21 (Location experiment—Uniform errors). Assume:

- $\Omega \subseteq \mathbb{R}$ (no need to specify prior)
- $S_1 = S_2 = \mathbb{R}$
- 1 observes $s_1 = \omega + \varepsilon_1$ for $\varepsilon_1 \sim U(-\rho_1, \rho_1)$
- 2 observes $s_2 = \omega + \varepsilon_2$ for $\varepsilon_2 \sim U(-\rho_2, \rho_2)$
- If $\rho_1 = 1/2$ and $\rho_2 = 1$, then 1 has better information: Conditional on ω, s_2 is distributed the same as $s_1 + \varepsilon$ where $\varepsilon = (-1/2, 0.5; 1/2, 0.5)$.
- However, it can be shown that if $\rho_1 < \rho_2 < 2\rho_1$, then it is *impossible* for 1 to replicate 2's signal distribution conditional on ω.
- Thus, ordering information by requiring that 1 can replicate 2's signal distribution is quite restrictive.

Example 6.21 demonstrates that our intuitions about when one signal is (Blackwell) more informative than another might be incorrect or misleading. In particular, by Theorem 6.14, if $\rho_1 < \rho_2 < 2\rho_1$ then there is some expected-utility maximizer who prefers signal σ_2 to σ_1 despite the smaller variance of the noise in signal σ_1.

However, Lehmann (1988) observed that if we further restrict the class of utility functions that individuals might have, then signal σ_1 in Example 6.21 will be preferred to σ_2 whenever $\rho_1 < \rho_2$. In other words, if we don't require that *all* expected-utility maximizers prefer σ_1 to σ_2, and instead only require this preference for a certain subclass of utility functions, then we are able to compare more signals. The so-called **Lehmann information order** therefore extends the Blackwell information order to permit the comparison of more information structures (signals). The interested reader is referred to the original work of Lehmann (1988), as well as the nice application of the Lehmann information order in Quah and Strulovici (2009).

6.4 Exercises

6.1 In this exercise, you are asked to prove the missing parts of the proof of Theorem 6.7, the result that characterizes different equivalent representations of $\sigma_1 : \Omega \to S_1$ being more informative than $\sigma_2 : \Omega \to S_2$.

(a) Prove (2) $\Rightarrow$ (1). That is, prove that if there exists a function $\gamma : S_1 \to S_2$ such that $\sigma_2 = \gamma \circ \sigma_1$, then the partition $\mathcal{E}_1$ generated by σ_1 is finer than the partition $\mathcal{E}_2$ generated by σ_2.

(b) Before proving the other missing part of the proof, first prove the following: If

$f : \Omega \to X$ and $g : \Omega \to Y$ satisfy

$$f(\omega) = f(\omega') \implies g(\omega) = g(\omega'), \quad \forall \omega, \omega' \in \Omega,$$

then there exists $h : X \to Y$ such that $g = h \circ f$. This will be useful for the next part of the problem, but also verifies the claim in the notes about the two equivalent ways of defining a feasible state-contingent action plan.

(c) Prove (1) $\Rightarrow$ (2). That is, prove that if the partition $\mathcal{E}_1$ generated by σ_1 is finer than the partition $\mathcal{E}_2$ generated by σ_2, then there exists a function $\gamma : S_1 \to S_2$ such that $\sigma_2 = \gamma \circ \sigma_1$. (Hint: Use part (b).)

6.2 Suppose there are two states, $\Omega = \{G, B\}$. In each of the following parts of this problem, you will be given a pair of signals $\sigma_1 : \Omega \to \triangle(S_1)$ and $\sigma_2 : \Omega \to \triangle(S_2)$, and you will be asked to determine which is Blackwell more informative than the other, or if the two signals are not ranked (that is, neither is Blackwell more informative than the other). If you claim that one of the signals is more informative than the other, describe the garbling (e.g., using a table) of that signal that yields the other signal. In other words, if you claim that σ_1 is more informative than σ_2, then construct a garbling $\gamma : S_1 \to \triangle(S_2)$ such that $\sigma_2 = \gamma \circ \sigma_1$. Or, if you claim that σ_2 is more informative than σ_1, then construct a garbling $\gamma : S_2 \to \triangle(S_1)$ such that $\sigma_1 = \gamma \circ s_2$. If you claim that neither is more informative than the other, then explain carefully (but briefly) why this is the case.

(a) Suppose $S_1 = \{g, b\}$ and $S_2 = \{g, b, n\}$, and suppose σ_1 and σ_2 are defined as follows:

$\sigma_1(s_1|\omega)$

	g	b
G	0.8	0.2
B	0.2	0.8

$\sigma_2(s_2|\omega)$

	g	b	n
G	0.6	0	0.4
B	0	0.6	0.4

(b) Suppose $S_1 = \{g, b\}$ and $S_2 = \{g, b, n\}$, and suppose σ_1 and σ_2 are defined as follows:

$\sigma_1(s_1|\omega)$

	g	b
G	0.6	0.4
B	0.4	0.6

$\sigma_2(s_2|\omega)$

	g	b	n
G	0.3	0.2	0.5
B	0.2	0.3	0.5

(c) Suppose $S_1 = S_2 = \{g, b\}$, and suppose σ_1 and σ_2 are defined as follows:

$\sigma_1(s_1|\omega)$

	g	b
G	0.8	0.2
B	0.4	0.6

$\sigma_2(s_2|\omega)$

	g	b
G	0.6	0.4
B	0.3	0.7

6.3 Suppose there are two states, $\Omega = \{G, B\}$. Consider three stochastic signals $\sigma_i : \Omega \to \triangle(S)$ for $i = 1, 2, 3$ where $S = \{g, b, n\}$. Each of these signals depends on a parameter $\eta_i \in [0, 1]$ and is defined as follows:

$$\begin{array}{c|c|c|c|}
\multicolumn{4}{c}{\sigma_i(s_i|\omega)} \\
\multicolumn{1}{c}{} & \multicolumn{1}{c}{g} & \multicolumn{1}{c}{b} & \multicolumn{1}{c}{n} \\
\cline{2-4}
G & \eta_i & 0 & 1-\eta_i \\
\cline{2-4}
B & 0 & \eta_i & 1-\eta_i \\
\cline{2-4}
\end{array}$$

Answer the following.

(a) What relationship between η_1 and η_2 ensures that σ_1 is Blackwell more informative than σ_2? Verify whatever relationship between η_1 and η_2 you claim is sufficient by constructing a garbling $\gamma : S \to \triangle(S)$ of σ_1 that gives σ_2. That is, provide the values of $\gamma(s_2|s_1)$ (e.g., in a table) and verify that applying this garbling to σ_1 yields σ_2. (Note: You are not required to prove that your asserted relationship between η_1 and η_2 is necessary for σ_1 to be more informative than σ_2; that is, you do not need to prove that σ_1 is *not* more informative than σ_2 when your condition is violated. However, finding the weakest possible sufficient condition and avoiding unnecessary restrictions will earn more points.)

(b) Consider a fourth stochastic signal that is obtained by observing the signal realizations of both σ_2 and σ_3 (and assume that the realizations of these two signals are independent once we condition on the state ω). Intuitively, if we think of these stochastic signals as the information held by different individuals, then individual 4 gets to observe the information of both individuals 2 and 3. Formally, $\sigma_4 : \Omega \to \triangle(S^2)$ gives the probability of signal realizations of the form $s_4 = (s_2, s_3) \in S^2$ for each $\omega \in \Omega$, and is defined by $\sigma_4((s_2, s_3)|\omega) = \sigma_2(s_2|\omega)\sigma_3(s_3|\omega)$ (we simply take the product of the probabilities for the two signal realizations since we assumed that they are independent once we condition on ω). If $\eta_1 = \eta_2 = \eta_3$, is σ_4 more informative than σ_1? If so, show there is a garbling $\gamma : S^2 \to \triangle(S)$ of σ_4 that gives σ_1; that is, provide the values of $\gamma(s_1|(s_2, s_3))$ and show that applying this garbling to σ_4 yields σ_1. If not, provide a careful explanation of why σ_4 is not more informative than σ_1.

(c) Consider again the stochastic signal σ_4 defined in the previous part of the question. However, now relax the assumption that $\eta_1 = \eta_2 = \eta_3$, and instead allow all three parameters take any values $\eta_1, \eta_2, \eta_3 \in [0, 1]$ (possibly all different or some the same—we are not assuming anything at this point). Given η_2 and η_3, determine the maximum value of η_1 (as a function of η_2 and η_3) such that σ_4 is more informative than σ_1. In particular, is it possible to have $\eta_1 > \eta_2$ and $\eta_1 > \eta_3$? For this maximum value of η_1, show there is a garbling $\gamma : S^2 \to \triangle(S)$ of σ_4 that gives σ_1. That is, provide the values of $\gamma(s_1|(s_2, s_3))$ and show that applying this garbling to σ_4 yields σ_1. (Note: You are not required to prove that your asserted maximum value of η_1 is the largest possible given η_2 and η_3; that is, you do not need to prove that σ_4 is *not* more informative than σ_1 when η_1 is above this value. However, finding the largest possible value of η_1 will earn more points.)

Chapter 7

Signals and Posterior Beliefs

Contents

References and Assigned Readings

Primary readings:

(1) Milgrom (1981) (optional).

Additional references:

(1) Blackwell (1951, 1953)
(2) Gollier (2001) (Chapter 24)

7.1 Distributions Over Posteriors

7.1.1 Modeling Information Using Distributions Over Posteriors

In this section, we again restrict attention to finite state and signal spaces. Given a prior P over Ω and a stochastic signal $\sigma : \Omega \to \triangle(S)$, the joint probability of state ω and signal realization s is

$$P_\sigma(\omega, s) = P(\omega)\sigma(s|\omega).$$

The unconditional probability of observing signal realization s is

$$P_\sigma(s) = \sum_{\omega \in \Omega} P_\sigma(\omega, s) = \sum_{\omega \in \Omega} P(\omega)\sigma(s|\omega).$$

The posterior probability distribution obtained by Bayesian updating after observing a signal realization s is given by

$$P_\sigma(\omega|s) = \frac{P_\sigma(\omega, s)}{P_\sigma(s)} = \frac{P(\omega)\sigma(s|\omega)}{\sum_{\omega'} P(\omega')\sigma(s|\omega')}$$

whenever $P_\sigma(s) > 0$.

Lemma 7.1. For a given prior P and a signal $\sigma : \Omega \to \triangle(S)$, Bayesian updating results in a unique distribution over posteriors $\mu \in \triangle(\triangle(\Omega))$. Moreover, μ satisfies the condition:

$$\int_{\triangle(\Omega)} p \, d\mu(p) = P. \tag{7.1}$$

Equation (7.1) is sometimes called the **Bayesian plausibility** condition. It states that beliefs are a Martingale: The expectation of posterior beliefs under the measure μ is precisely the prior.

Proof. Given a prior P and a signal $\sigma : \Omega \to \triangle(S)$, we can construct a distribution over posteriors as follows: First, let

$$S^* = \{s \in S : P_\sigma(s) > 0\}$$

denote the set of signal realizations that occur with positive probability. For each $s \in S^*$, let p_s denote the posterior probability distribution obtained through Bayesian updating following this signal realization. That is, p_s is defined by

$$p_s(\omega) = P_\sigma(\omega|s) = \frac{P(\omega)\sigma(s|\omega)}{\sum_{\omega'} P(\omega')\sigma(s|\omega')} \quad \forall \omega \in \Omega.$$

Define a distribution over posteriors $\mu \in \triangle(\triangle(\Omega))$ by

$$\mu(p) = \sum_{s \in S^*} P_\sigma(s)\mathbf{1}_{p_s}(p)$$

for any $p \in \triangle(\Omega)$, where $\mathbf{1}_{p_s}(p)$ is the indicator function that takes value 1 if $p_s = p$ and takes value zero otherwise. In words, the probability that μ assigns to the posterior p is the probability under P_σ of observing a signal realization s that results in posterior belief p. Since multiple signal realizations could result in the same posterior belief, we must sum over all s that give posterior $p_s = p$ in order to calculate that total probability of p. Note that since S^* is finite, μ has finite support.[1] Finally, we verify that Equation (7.1) holds. For any $\omega \in \Omega$,[2]

$$\begin{aligned}
\sum_{p \in \triangle(\Omega)} p(\omega)\mu(p) &= \sum_{p \in \triangle(\Omega)} p(\omega) \sum_{s \in S^*} P_\sigma(s)\mathbf{1}_{p_s}(p) \\
&= \sum_{s \in S^*} \sum_{p \in \triangle(\Omega)} p(\omega) P_\sigma(s)\mathbf{1}_{p_s}(p) = \sum_{s \in S^*} p_s(\omega)P_\sigma(s) \\
&= \sum_{s \in S^*} P_\sigma(\omega|s)P_\sigma(s) = \sum_{s \in S^*} P_\sigma(\omega, s) = P(\omega).
\end{aligned}$$

This completes the proof. □

As the proof of Lemma 7.1 shows, finiteness of the set S implies that μ has finite support, meaning that only finitely many posteriors p are assigned positive probability. This allows us to write expectations over posteriors as sums rather than integrals.

The following result shows that there is a convenient expression for ex-ante expected utility in terms of the distribution over posteriors generated by a signal.

Lemma 7.2. Fix any set of actions A, utility function $u : A \times \Omega \to \mathbb{R}$, and prior P. Then,

[1] In other words, if we enumerate the finite set of signal realizations that occur with positive probability as $S^* = \{s_1, \ldots, s_n\}$ for some $n \in \mathbb{N}$, then we have $\mu \in \triangle(\{p_{s_1}, \ldots, p_{s_n}\})$.

[2] Although the set $\triangle(\Omega)$ is infinite, we can still write the sum over $p \in \triangle(\Omega)$ in what follows, since these terms are equal to zero for all but finitely many p given that μ has finite support.

the **ex-ante expected utility** from a signal σ can be expressed in terms of the resulting distribution over posteriors μ as follows:

$$\sum_{p\in\triangle(\Omega)} \mu(p)\left(\max_{a\in A}\sum_{\omega\in\Omega} u(a,\omega)p(\omega)\right).$$

Proof. This result follows from simple algebraic manipulations of the formula for ex-ante expected utility:

$$\begin{aligned}
&\max_{\alpha:S\to\triangle(A)}\sum_{\omega\in\Omega}\sum_{s\in S}\sum_{a\in A} u(a,\omega)\alpha(a|s)\sigma(s|\omega)P(\omega)\\
&=\max_{\alpha:S\to\triangle(A)}\sum_{\omega\in\Omega}\sum_{s\in S^*}\sum_{a\in A} u(a,\omega)\alpha(a|s)P_\sigma(\omega|s)P_\sigma(s)\\
&=\sum_{s\in S^*} P_\sigma(s)\max_{\beta\in\triangle(A)}\sum_{\omega\in\Omega}\sum_{a\in A} u(a,\omega)\beta(a)P_\sigma(\omega|s)\\
&=\sum_{s\in S^*} P_\sigma(s)\max_{a\in A}\sum_{\omega\in\Omega} u(a,\omega)p_s(\omega)\\
&=\sum_{p\in\triangle(\Omega)} \mu(p)\max_{a\in A}\sum_{\omega\in\Omega} u(a,\omega)p(\omega).
\end{aligned}$$

Note that the second-to-last equality switches from maximizing over lotteries $\beta\in\triangle(A)$ to maximizing over deterministic actions $a\in A$. This equality holds because of the linearity of the objective function in β. □

The following result is in some sense a converse to Lemma 7.1. Given a distribution over posteriors μ, we can construct a prior P, signal space S, and signal σ that together generate μ. Although this claim is true in general, we will focus on the case of measures μ with finite support since we are restricting attention to finite signal spaces S.

Lemma 7.3. Given a distribution over posteriors $\mu\in\triangle(\triangle(\Omega))$ with finite support, there exists a prior P and signal $\sigma:\Omega\to\triangle(S)$ that generate μ by Bayesian updating.

Proof. Since μ has finite support, we can enumerate the elements of its support as $\{p_1,\dots,p_n\}$ for some n. Let the signal space be $S=\{1,\dots,n\}$. Define a prior P by

$$P(\omega)=\sum_{s\in S} p_s(\omega)\mu(p_s)$$

and define $\sigma:\Omega\to\triangle(S)$ by

$$\sigma(s|\omega)=\frac{p_s(\omega)\mu(p_s)}{\sum_{s'\in S} p_{s'}(\omega)\mu(p_{s'})}.$$

Given these definitions of P and σ, we have

$$P_\sigma(\omega, s) = \sigma(s|\omega)P(\omega) = p_s(\omega)\mu(p_s).$$

This implies that the probability of observing signal realization s is

$$P_\sigma(s) = \sum_{\omega\in\Omega} P_\sigma(\omega, s) = \sum_{\omega\in\Omega} p_s(\omega)\mu(p_s) = \mu(p_s)$$

and the posterior belief following signal realization s is

$$P_\sigma(\omega|s) = \frac{P_\sigma(\omega, s)}{P_\sigma(s)} = p_s(\omega).$$

Thus, the distribution over posteriors generated by P and σ is precisely μ. $\square$

Note that multiple signals σ could lead to the same distribution over posteriors μ. In this sense, distributions over posteriors serve as a *canonical* representation of information structures: They capture all of the payoff-relevant information associated with the information structure. Blackwell (1951, 1953) referred to the induced distribution over posteriors as the **standard measure** of an experiment. Not surprisingly, distributions over posteriors have become the prevailing modeling device for information structures in economics recently, for example, in the literatures on information design and Bayesian persuasion.

7.1.2 Comparing Information Using Distributions Over Posteriors

Definition 7.4. Given two distributions over posteriors $\mu_1, \mu_2 \in \triangle(\triangle(\Omega))$, we say that μ_1 dominates μ_2 in the **convex order** if

$$\int_{\triangle(\Omega)} \varphi(p)\, d\mu_1(p) \geq \int_{\triangle(\Omega)} \varphi(p)\, d\mu_2(p)$$

for all convex functions $\varphi : \triangle(\triangle(\Omega)) \to \mathbb{R}$.

Of course, given our focus on distributions over posteriors with finite support in this section, we can write this condition using sums instead of integrals:

$$\sum_{p\in\triangle(\Omega)} \varphi(p)\, \mu_1(p) \geq \sum_{p\in\triangle(\Omega)} \varphi(p)\, \mu_2(p).$$

Intuitively, the convex order relates to the ex-ante expected utility in a decision problem (see the formula in Lemma 7.2) since

$$\varphi(p) = \max_{a\in A} \sum_{\omega\in\Omega} u(a, \omega)p(\omega)$$

is a convex function of p.

The following result adds a fourth condition to our previous characterization theorem for the Blackwell information order. We only state the equivalence of our new condition with the garbling condition on signals, but clearly via the equivalences established in Theorem 6.14, condition (2) below is equivalent to all of the conditions listed in that theorem.

Theorem 7.5 (Blackwell (1951, 1953)). Suppose $\sigma_1 : \Omega \to \triangle(S_1)$ and $\sigma_2 : \Omega \to \triangle(S_2)$ are two stochastic signals and fix any prior P with full support (i.e., $P(\omega) > 0$ for all $\omega \in \Omega$). The following are equivalent:

(1) σ_2 is a garbling of σ_1.
(2) If μ_1 and μ_2 are the distributions over posteriors generated by P and the signals σ_1 and σ_2, respectively, then μ_1 dominates μ_2 in the convex order.

7.2 MLR Property of Signals and Posteriors

In this section, we explore the implications of the monotone likelihood ratio property of signals. Note that this section is *not* about comparing information. Rather it is focused connecting properties of a single signal with properties of the resulting posterior beliefs conditional on signal realizations.

Suppose $\Omega, S \subseteq \mathbb{R}$. Thinking of a stochastic signal $\sigma : \Omega \to \triangle(S)$ as a family of probability distributions $\{\sigma(\cdot|\omega)\}_{\omega\in\Omega}$ over S that is indexed by the state ω, the MLR property is defined just as in the previous chapters.

Definition 7.6. A signal $\sigma : \Omega \to \triangle(S)$ has the **monotone likelihood ratio property** if $\omega' > \omega$ implies $\sigma(\cdot|\omega') \geq_{MLR} \sigma(\cdot|\omega)$. More explicitly, $\omega' > \omega$ and $s' > s$ implies

$$\sigma(s'|\omega')\sigma(s|\omega) \geq \sigma(s|\omega')\sigma(s'|\omega). \tag{7.2}$$

If all of the terms in Equation (7.2) are strictly positive, it can of course be written as

$$\frac{\sigma(s'|\omega')}{\sigma(s'|\omega)} \geq \frac{\sigma(s|\omega')}{\sigma(s|\omega)}.$$

Lemma 7.7. Suppose $\Omega, S \subseteq \mathbb{R}$ and the stochastic signal $\sigma : \Omega \to \triangle(S)$ has the monotone likelihood ratio property. Then, for any prior P over Ω, the posterior probability distributions over Ω have the monotone likelihood ratio property in the signal, that is, if $s' > s$ and $P_\sigma(s), P_\sigma(s') > 0$ then $P_\sigma(\cdot|s') \geq_{MLR} P_\sigma(\cdot|s)$.[3]

[3] We are restricting attentions to signal realizations $s, s' \in S$ that are reached with positive probability to

Proof. Fix any $\omega' > \omega$ and $s' > s$. Multiplying both sides of Equation (7.2) by $P(\omega')P(\omega)$ gives

$$P_\sigma(s', \omega')P_\sigma(s, \omega) \geq P_\sigma(s, \omega')P_\sigma(s', \omega).$$

Dividing both sides by $P_\sigma(s')P_\sigma(s)$ (recall that we are restricting attention to s, s' for which these are strictly positive) gives

$$P_\sigma(\omega'|s')P_\sigma(\omega|s) \geq P_\sigma(\omega'|s)P_\sigma(\omega|s').$$

Thus $\{P_\sigma(\cdot|s)\}_{s \in S}$ has the MLR property. $\square$

Since $P_\sigma(\cdot|s') \geq_{MLR} P_\sigma(\cdot|s)$ implies $P_\sigma(\cdot|s') \geq_{FOSD} P_\sigma(\cdot|s)$ (see Lemma 4.14), this result tells us in particular that higher signal realizations lead to first-order stochastic dominance increases in the posterior distribution over states, for every possible prior P. It turns out that the converse is also true.

Theorem 7.8 (Milgrom (1981)). Suppose $\Omega, S \subseteq \mathbb{R}$, and consider a stochastic signal $\sigma : \Omega \to \triangle(S)$. Then the following are equivalent:

(1) σ has the monotone likelihood ratio property.
(2) For any prior P, if $s' > s$ and $P_\sigma(s), P_\sigma(s') > 0$ then $P_\sigma(\cdot|s') \geq_{FOSD} P_\sigma(\cdot|s)$.

Proof. (1) $\Rightarrow$ (2): As observed already, this follows from Lemmas 7.7 and 4.14.

(2) $\Rightarrow$ (1): Fix any $\omega' > \omega$ and $s' > s$. Take the prior P such that $P(\omega) = P(\omega') = 0.5$. If $P_\sigma(s) = 0$ then it must be that $\sigma(s|\omega) = \sigma(s|\omega') = 0$, in which case Equation (7.2) is trivially satisfied. Likewise, if $P_\sigma(s') = 0$ then $\sigma(s'|\omega) = \sigma(s'|\omega') = 0$. The remaining case is $P_\sigma(s), P_\sigma(s') > 0$. In this case, since the support of $P_\sigma(\cdot|s)$ and $P_\sigma(\cdot|s')$ contains only two states, ω and ω', condition (2) requires that $P_\sigma(\omega'|s') \geq P_\sigma(\omega'|s)$ and $P_\sigma(\omega|s') \leq P_\sigma(\omega|s)$. Thus

$$P_\sigma(\omega'|s')P_\sigma(\omega|s) \geq P_\sigma(\omega'|s)P_\sigma(\omega|s').$$

Multiplying both sides by $P_\sigma(s)P_\sigma(s')$ gives

$$P_\sigma(\omega', s')P_\sigma(\omega, s) \geq P_\sigma(\omega', s)P_\sigma(\omega, s').$$

Dividing both sides by $P(\omega)P(\omega') = 1/4$ gives

$$\sigma(s'|\omega')\sigma(s|\omega) \geq \sigma(s|\omega')\sigma(s'|\omega).$$

This completes the proof. $\square$

ensure that the posteriors conditional these signal realizations are uniquely defined.

Bibliography

Athey, S. (2002): "Monotone Comparative Statics under Uncertainty," *Quarterly Journal of Economics*, 117, 187–223. [90,96,99]

Blackwell, D. (1951): "Comparisons of Experiments," *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability,* University of California Press, 93–102. [79,112,120,132,135,136,141]

Blackwell, D. (1953): "Equivalent Comparisons of Experiments," *The Annals of Mathematical Statistics*, 24, 265–272. [79,112,120,132,135,136,141]

Bulow, J., J. Geanakoplos, and P. Klemperer (1985): "Multimarket Oligopoly: Strategic Substitutes and Complements," *Journal of Political Economy*, 93, 488–511. [61]

de Oliveira, H. (2018): "Blackwell's Informativeness Theorem Using Diagrams," *Games and Economic Behavior*, 109, 126–131. [112,121]

Gollier, C. (2001): *The Economics of Risk and Time*, Cambridge, MA: MIT Press. [68,90,95,96,112,132]

Karlin, S. (1968): *Total Positivity: Volume I*, Stanford, CA: Stanford University Press. [96]

Kreps, D. (2013): *Microeconomic Foundations I: Choice and Competitive Markets*, Princeton, NJ: Princeton University Press. [83]

Lehmann, E. L. (1988): "Comparing Location Experiments," *Annals of Statistics*, 16, 521–533. [127]

Machina, M. J., and J. W. Pratt (1997): "Increasing Risk: Some Direct Constructions," *Journal of Risk and Uncertainty*, 14, 103–127. [68,77]

Mas-Colell, A., M. D. Whinston, and J. R. Green (1995): *Microeconomic Theory*, Oxford University Press. [68,83,112]

Milgrom, P. (1981): "Good News and Bad News: Representation Theorems and Applications," *Bell Journal of Economics*, 12, 380–391. [132,137,141]

Milgrom, P., and J. Roberts (1990): "Rationalizability, Learning, and Equilibrium in Games with Strategic Complementarities," *Econometrica*, 58, 1255–1277. [42,55,56,141]

Milgrom, P., and J. Roberts (1996): "The LeChatelier Principle," *American Economic Review*, 86, 173–179. [22,33,34,35,141]

Milgrom, P., and C. Shannon (1994): "Monotone Comparative Statics," *Econometrica*, 62, 157–180. [2,12,22,24,30,44,55,141]

Müller, A., and D. Stoyan (2002): *Comparison Methods for Stochastic Models and Risks*, West Sussex, England: John Wiley & Sons. [79]

Pratt, J. W. (1964): "Risk Aversion in the Small and in the Large," *Econometrica*, 32, 122–136. [68,83]

Quah, J., and B. Strulovici (2009): "Comparative Statics, Informativeness, and the Interval Dominance Order," *Econometrica*, 77, 1949–1992. [26,127]

Rothschild, M., and J. E. Stiglitz (1970): "Increasing Risk I: A Definition," *Journal of Economic Theory*, 2, 225–243. [68,78,79]

Rothschild, M., and J. E. Stiglitz (1971): "Increasing Risk II: Its Economic Consequences," *Journal of Economic Theory*, 3, 66–84. [90]

Rothschild, M., and J. E. Stiglitz (1972): "Addendum to 'Increasing Risk I: A Definition'," *Journal of Economic Theory*, 5, 306. [68,79]

Shaked, M., and J. G. Shanthikumar (2007): *Stochastic Orders*, New York, NY: Springer. [77,79]

Shannon, C. (1995): "Weak and Strong Monotone Comparative Statics," *Economic Theory*, 5, 209–227. [96]

Topkis, D. M. (1978): "Minimizing a Submodular Function on a Lattice," *Operations Research*, 26, 305–321. [7,11,141]

Topkis, D. M. (1979): "Equilibrium Points in Nonzero-Sum n-Person Submodular Games," *Siam Journal of Control and Optimization*, 17, 773–787. [44,48,141]

Topkis, D. M. (1998): *Supermodularity and Complementarity*, Princeton, NJ: Princeton University Press. [2,22,42]

Vives, X. (1990): "Nash Equilibrium with Strategic Complementarities," *Journal of Mathematical Economics*, 19, 305–321. [42,56]

Index