Course Handouts

ECON 8101

MICROECONOMIC THEORY

Jan Werner

University of Minnesota

FALL SEMESTER 2007

PART I: Producer Theory

1. Production Set

Production set is a subset Y of commodity space $I\!R^L$, where L is the number of commodities. Vectors in Y represent **production plans** that are technologically feasible.

Negative coordinates of production plan $y = (y_1, \ldots, y_L) \in Y$ are understood as input quantities; positive coordinates of y are output quantities.

Production plan $y \in Y$ is **efficient** if there is no alternative production plan $y' \in Y$, $y' \neq y$, such that $y' \geq y$.

Example (*Activity analysis*):

If two activities $a^1, a^2 \in I\!R^L$ can be combined together at arbitrary scale, then the production set is $Y = \{y \in I\!R^L : y = \lambda_1 a^1 + \lambda_2 a^2,\ \lambda_1 \geq 0, \lambda_2 \geq 0\}$.

Some **properties of production sets**:

(i) Y closed; $0 \in Y$.

(ii) *no free production:* $Y \cap I\!R^L_+ = \{0\}$.

(iii) *free disposal:* $Y - I\!R^L_+ \subset Y$.

(iv) Y convex,

- Property (i) will be assumed throughout.

A convenient specification of a production set is in the form

$$Y = \{y \in I\!R^L : T(y) \leq 0\} \tag{1}$$

for some function $T : I\!R^L \to I\!R$, called **transformation function**. Typically, function T is increasing, continuous, and such that $T(0) = 0$. Such specification permits the use of marginal rates of transformation $\frac{\partial T}{\partial y_i} / \frac{\partial T}{\partial y_j}$.

Production function:

Often in applied work and in examples, production technology is specified by a **production function**. In the simple case of single output, production function is $f : I\!R^n_+ \to I\!R_+$ that associates a quantity of single output with a vector of some n inputs. We write $f(x) = z$, where $x = (x_1, \ldots, x_n)$ is a vector of inputs (here with positive sign!).

Examples: Cobb-Douglas, Leontief, CES, etc.

Some **properties of production functions:**

(i) $f(0) = 0$; f continuous (or differentiable) function.

(ii) f concave or quasi-concave.

Production function f gives rise to production set Y_f given by

$$Y_f = \{(x, z) \in I\!R^{n+1} : \; x \leq 0, \; 0 \leq z \leq f(-x)\}. \tag{2}$$

2. Returns to Scale in Production

Properties of returns to scale for production set are defined as follows:

constant – if $y \in Y$, then $\lambda y \in Y$ for every $\lambda \geq 0$,

nonincreasing – if $y \in Y$, then $\lambda y \in Y$ for every $0 \leq \lambda \leq 1$,

nondecreasing – if $y \in Y$, then $\lambda y \in Y$ for every $\lambda \geq 1$,

Actually, returns to scale can be more crisply defined for production function. These definitions are

constant: $f(\lambda x) = \lambda f(x)$, for every $\lambda \geq 0$ and $x \geq 0$.

decreasing: $f(\lambda x) < \lambda f(x)$, for every $\lambda > 1$ and $x \geq 0$ such that $f(x) \neq 0$.

increasing: $f(\lambda x) > \lambda f(x)$, for every $\lambda > 1$ and $x \geq 0$ such that $f(x) \neq 0$.

One can show (Exercise) that constant, decreasing or increasing returns to scale for f imply that the production set Y_f of (2) exhibits constant, nonincreasing or nondecreasing returns to scale, respectively.

3. Profit Maximization

Profit maximization at price vector $p \in I\!R^L$ is

$$\max\{py : y \in Y\}$$

The solutions (there could be many) are the **supply** of the firm at p, denoted by $s^*(p)$. We can write

$$s^*(p) = \{y^* \in Y : py^* \geq py, \ \forall y \in Y\}. \tag{3}$$

The (maximum) **profit** is

$$\pi^*(p) = \max_{y \in Y} py. \tag{4}$$

π^* is a function of p while s^* is, in general, a correspondence.

If supply s^* is a differentiable function, then the $L \times L$-matrix $Ds^*(p)$ is called **the supply substitution matrix.**

The set of price vectors for which profit function π^* is well defined (*domain*) is a convex set in $I\!R^L$. The domain of s^* is a subset of the domain of π^*.

If production set Y is compact, then domains of π^* and s^* are $I\!R^L$ - all price vectors. If production set Y exhibits constant returns to scale, then the domains of π^* and s^* are the *polar cone* Y°, that is the set

$$Y^\circ = \{p \in I\!R^L : py \leq 0, \ \forall y \in Y\}. \tag{5}$$

4. Supply and Profit

Fundamental properties of the supply and the profit function of a profit-maximizing firm are:

Theorem 4.1: *Suppose that Y is closed. Then the following hold for π^* and s^* on their respective domains:*

(i) π^* is a continuous function;

(ii) π^* is homogeneous of deg. 1;

(iii) π^* is a convex function;

(iv) s^* is homogeneous of deg. 0;

(v) If π^* is differentiable at p (this holds iff s is single-valued at p), then

$$D\pi^*(p) = s^*(p).$$

Proof: (i) if Y is compact, then this follows from the Maximum Theorem; the general case is discussed in class; (ii) and (iv) are easy; (iii) do this one on your own! (v) see MWG, Section 5.C, pg. 138.

Using (iii) and (v) of Theorem 4.1, we obtain

Corollary 4.2: *If π^* is twice-differentiable, then*

$$D^2\pi^*(p) = Ds^*(p) \tag{6}$$

. The substitution matrix $Ds^*(p)$ is positive semi-definite and symmetric.

Corollary 4.2 implies the following **comparative statics** property of supply:

$$\frac{\partial s_i^*}{\partial p_i} \geq 0. \tag{7}$$

Some **extra properties** of supply and profit of profit-maximizing firm:

Proposition 4.3:

(i) if Y exhibits constant returns to scale, then $\pi^*(p) = 0$ whenever well-defined.

(ii) if Y is convex, then $s^*(p)$ is a convex set.

(iii) if Y is compact, then correspondence s^* is upper hemi-continuous.

Proof: (i) and (ii) left as exercises, (iii) follows from the Maximum Theorem.

Supply and profit for production functions:

Profit maximization for production function is typically written as

$$\max_{x_1 \geq 0, \ldots, x_n \geq 0} [qf(x_1, \ldots, x_n) - \sum_{i=1}^{n} w_i x_i]$$

or, for short,

$$\max_{x \geq 0} [qf(x) - wx],$$

where q denotes the price of (single) output and $w = (w_1, \ldots, w_n)$ input prices.

The solutions (again, there could be many) are the **input (or factor) demand** at prices (w, q), denoted by $x^*(w, q)$.

The (maximum) **profit** is $\pi^*(w, q) = \max_{x \geq 0} [qf(x) - wx]$.

There is also the profit-maximizing **output supply** $z^*(w, q) = f(x^*(w, q))$.

If $q > 0$, then these solutions for f coincide with the supply and profit of the production set Y_f given by (2). More precisely, supply s^* of Y_f is $(-x^*, z^*)$ and the maximum profit π^* of f and of Y_f are the same function.

Hence, the properties of Theorem 4.1 and Corollary 4.2 apply to production function.

5. Profit-rationalizability

Consider a function π that assigns profit to each price vector p in its domain which is a convex subset of $I\!R^L$. Call π a profit function, but it is not known whether or not π results from maximizing profit on some production set, that is, whether π is a maximum profit function. Production set Y **profit-rationalizes** profit function π if $\pi(p) = \max\{py : y \in Y\}$ for every p. Properties *(i)*, *(ii)*, and *(iii)* of Theorem 4.1 turn out to be sufficient for profit-rationalizability.

Theorem 5.1: *If π is (i) continuous, (ii) homogeneous of deg 1, and (iii) convex, then there exists a closed and convex set Y that profit-rationalizes π (with possible exception of some price vectors on the relative boundary of the domain of π).*

Proof – This theorem is an application of Corollary 13.2.1 in Rockafellar's "Convex Analysis." The set Y that profit-rationalizes function π is

$$Y = \{y \in I\!R^L : py \le \pi(p),\ \forall p\}.$$

Chapter 13 in Rocka's book is about the so-called support functions, see MWG, Section 3.F. Technically speaking, maximum profit function π^* is the support function of the production set.

Consider a function ψ that assigns a production plan to each price vector p in its domain which is a subset of $I\!R^L$. Call ψ a supply function, but again it is not known whether or not ψ results from maximizing profit on some production set. The profit function associated with supply ψ is $\pi(p) = p\psi(p)$. If ψ is homogeneous of deg 0 (property *(iv)* of Theorem 4.1), then so defined π is homogeneous of deg 1. Further, if the substitution matrix $D\psi(p)$ is symmetric and positive semi-definite, then $D^2\pi(p) = D\psi(p)$ and π is convex (*(v)* of Theorem 4.1). Then Theorem 5.1 implies that there exists Y that profit-rationalizes π and ψ.

6. Weak Axiom of Profit Maximization

Suppose that we have several observations of prices and production plans of a firm. They are

$$\begin{array}{ccc} y^1 & \text{at} & p^1, \\ \cdots & & \cdots, \\ y^T & \text{at} & p^T. \end{array}$$

If the firm maximizes its profit, it must hold that

$$p^t y^t \geq p^t y^s, \tag{8}$$

for all $s = 1, \ldots, T$, for each t.

Property (8) is called the **Weak Axiom of Profit Maximization**.

Production set Y **profit-rationalizes** observations $(p^1, y^1), \ldots, (p^T, y^T)$ if $y^t \in Y$ and $p^t y^t = \max\{p^t y : y \in Y\}$ for every t.

Proposition 6.1: *Observations $(p^1, y^1), \ldots, (p^T, y^T)$ satisfy WAPM if and only if there exists a closed, convex production set Y that profit-rationalizes these observations.*

Write WPAM twice as

$$p^s y^s \geq p^s y^t$$

and

$$p^t y^t \geq p^t y^s,$$

and add side by side. We obtain

$$[p^t - p^s][y^t - y^s] \geq 0.$$

One can write this as

$$\Delta p \Delta y \geq 0 \tag{9}$$

Property (9) can be viewed as algebraic comparative statics of profit maximization. For instance, if all prices except for the price of good i are the same at two observations s and t, i.e., $p_j^t = p_j^s$ for $j \neq i$, and if good i is an output good, then (9) indicates that the output supply of a profit maximizing firm increases when the price of the output increases. The calculus comparative-statics expression for this is $\frac{\partial s_i^*}{\partial p_i} \geq 0$.

PART II: Consumer Theory

7. Preferences and Utility Functions

Consumption set is a subset $X \subset \mathcal{R}^L$. Vectors in X represent consumption bundles that the consumer considers possible for consumption.

Often, it is assumed that consumption set X is closed and convex, or more specifically that $X = \mathbb{R}^L_+$.

The consumer's preferences over commodity bundles in X are specified by a **preference relation** $\succeq$.

Properties that a preference relation may have:

(i) reflexive, transitive and complete,

(ii) continuous,

(iii) nonsatiated, or locally nonsatiated,

(iv) increasing, or strictly increasing (also called weakly or strictly monotone),

(v) convex, or strictly convex.

Other special properties: homothetic, quasi-linear, etc.

Examples of preferences: lexicographic; Leontief; etc.

Function $u : X \to \mathcal{R}$ is a **utility representation** of $\succeq$ if, for every $x, x' \in X$,

$$u(x) \geq u(x') \quad \text{if and only if } x \succeq x'. \tag{10}$$

Theorem 7.1: *If preference relation $\succeq$ on X is complete, reflexive, transitive, and continuous, then it has a (continuous) utility representation.*

Proof: See Hildenbrand and Kirman (1976). An easy proof is available if two additional assumptions are imposed: $X = \mathcal{R}^L_+$, and $\succeq$ strictly increasing. This proof can be found in MWG and in Varian.

8. Utility Maximization

The problem of utility maximization for a price vector $p \in \mathcal{R}^L_+$ and an income $w > 0$ is written as

$$\text{maximize} \quad u(x) \tag{11}$$

$$\text{subject to} \quad px \leq w \quad \text{and} \quad x \geq 0.$$

Solution is denoted by $x^*(p, w)$ – it is the **demand** at prices p and income w. Demand $x^*(p, w)$ is often called **Walrasian** or **Marshallian**.

$u^*(p, w) \equiv u(x^*(p, w))$ is the **indirect utility function.**

9. Expenditure Minimization

The expenditure minimization problem for $p \in \mathcal{R}^L_+$ and utility level $\bar{u}$ in the image of u is:

$$\text{minimize} \quad px \tag{12}$$

$$\text{subject to} \quad u(x) \geq \bar{u} \quad \text{and} \quad x \geq 0.$$

Solution is $h(p, \bar{u})$ – **Hicksian demand** correspondence, or function whenever single-valued.

Also $e(p, \bar{u}) \equiv ph(p, \bar{u})$ is the **expenditure function.**

The fundamental properties of Hicksian demand and expenditure function are:

Theorem 9.1: *Suppose that u is a continuous and locally non-satiated utility function on $X = I\!R^L_+$. Then, on the domain of strictly positive prices,*

(i) e is homogeneous of deg. 1 in prices;

(ii) e is a concave function of prices;

(iii) h is homogeneous of deg. 0 in prices.

(iv) If e is differentiable at $(p, \bar{u})$ and h is single-valued at $(p, \bar{u})$, then

$$D_p e(p, \bar{u}) = h(p, \bar{u}). \tag{13}$$

Using (ii) and (iv) of Theorem 9.1, we obtain

Corollary 9.2: *If e is twice-differentiable with respect to prices, then $D^2_p e(p, \bar{u}) = D_p h(p, \bar{u})$. The matrix $D_p h(p, \bar{u})$ is negative semi-definite and symmetric.*

Corollary 9.2 implies the following **comparative statics** property of Hicksian demand:

$$\frac{\partial h_i}{\partial p_i} \leq 0. \tag{14}$$

Remark: The matrix $D_p h(p, \bar{u})$ is singular. This is so because $p D_p h(p, \bar{u}) = 0$ as follows from (iii).

Digression on Cost Minimization.

The problem of cost minimization for a producer with production function is formally equivalent to expenditure minization.

Using the setup as in Section 4 (page 7), the problem of minimizing cost is written as

$$\text{minimize} \quad wx \tag{15}$$

$$\text{subject to} \quad f(x) \geq z \quad \text{and} \quad x \geq 0,$$

where $w = (w_1, \ldots, w_n)$ is a vector of input prices (*not income*!).

Solution is $x^*(w, z)$ – **factor demand** correspondence, or function whenever single-valued. Also $C^*(w, z) \equiv wx^*(w, z)$ is the **cost function.**

Theorem 9.1 when applied to cost minimization (15) says that the cost function C^* is a concave and homogeneous of deg 1 function of input prices. Further

$$D_w C^*(w, z) = x^*(w, z), \tag{16}$$

and

$$\frac{\partial x_i^*}{\partial w_i} \leq 0. \tag{17}$$

Results of Theorem 9.1 say nothing about how the cost function C^* depends on the output quantity z. This is clearly an interesting question in the context of cost minimization. We will have more to say about this in Part III.

10. Walrasian Demand and Hicksian Demand

Let $h(p, \bar{u})$ be the Hicksian demand and $x^*(p, w)$ be the Walrasian demand correspondences of utility function u on consumption set $X = \mathcal{R}^L_+$. Let $w > 0$, $\bar{u} > u(0)$ and $p >> 0$.

Proposition 10.1: *If u is continuous and locally non-satiated, then*

$$h(p, \bar{u}) = x^*(p, e(p, \bar{u})), \tag{18}$$

and

$$x^*(p, w) = h(p, u^*(p, w)). \tag{19}$$

Proof (Outline): We first have the following

Lemma 10.2: (1) If u is locally non-satiated, then $px^*(p, w) = w$.

(2) If u is continuous, then $u(h(p, \bar{u})) = \bar{u}$.

Step 1: Next we prove the following two relations:

(i') $h(p, \bar{u}) \subset x^*(p, e(p, \bar{u}))$

(ii') $x^*(p, w) \subset h(p, u^*(p, w))$

Step 2: From (i') it follows that $u^*(p, e(p, \bar{u})) = \bar{u}$. From (ii') it follows that $e(p, u^*(p, w)) = w$.

Step 3: Since $u^*(p, e(p, \bar{u})) = \bar{u}$, relation reverse to (i') follows from (ii'). Similarly, relation reverse to (ii') follows from (i') and $e(p, u^*(p, w)) = w$.

11. The Slutsky Equation and the Slutsky Matrix

Suppose that (18) holds and h and x^* are single-valued and differentiable. It follows that

$$D_p h(p,\bar{u}) = D_p x^*(p,w) + D_w x^*(p,w) \cdot x^*(p,w) \tag{20}$$

where $w = e(p,\bar{u})$, or equivalently $\bar{u} = u^*(p,w)$

More specifically (and rearranging)

$$\frac{\partial x_l^*(p,w)}{\partial p_k} = \frac{\partial h_l(p,\bar{u})}{\partial p_k} - \frac{\partial x_l^*(p,w)}{\partial w} x_k^*(p,w), \tag{21}$$

where $\bar{u} = u^*(p,w)$.

Equation (21) is the **Slutsky equation**. It provides decomposition of the effect of change in price of good k on Walrasian demand for good l into the **pure substitution effect** and the **income effect**.

Define the $L \times L$ matrix $S = [s_{kl}]$ by

$$s_{kl} = \frac{\partial x_l^*(p,w)}{\partial p_k} + \frac{\partial x_l^*(p,w)}{\partial w} x_k^*(p,w). \tag{22}$$

Matrix S is the **Slutsky matrix** associated with Walrasian demand x^*. It follows from (20) that

$$S(p,w) = D_p h(p,\bar{u}),$$

for $\bar{u} = u^*(p,w)$. Corollary 9.1 implies that S **is negative semi-definite, symmetric and satisfies** $S(p,w)p = 0$**.** Negative semi-definiteness of Slutsky matrix S is the comparative statics of Walrasian demand.

12. Integrability

We found in Section 11 that Walrasian demand function $x^*(p, w)$ of a utility-maximizing consumer with l.n.s. utility function necessarily has the following three properties: (1) homogeneous of deg. 0, (2) negative semi-definite and symmetric Slutsky matrix, (3) budget equation $px^*(p, w) = w$.

Question: Are these *all* properties of Walrasian demand functions?

One way to answer this question is to verify whether, for every demand function d that satisfies (1-3), a utility function u can be found such that the Walrasian demand function of u is the function d.

The answer is **yes**.

Theorem 12.1: *Let $d : \mathcal{R}^L_{++} \times \mathcal{R}_+ \to \mathcal{R}^L_+$ be a C^1 demand function such that*

(1) d is homogeneous of deg. 0,

(2) the Slutsky matrix associated with d is negative semi-definite and symmetric,

(3) $pd(p, w) = w$.

Then there exists a strictly increasing, strictly quasi-concave utility function u such that d is the Walrasian demand of utility function u.

Proof: see MWG, Section 3.H.

13. Revealed Preference

Suppose that we have several observations of price vectors and consumption plans of a consumer. They are

$$\begin{array}{ccc} x^1 & \text{at} & p^1, \\ \ldots & & \ldots, \\ x^T & \text{at} & p^T. \end{array}$$

We assume that $x^t \in \mathcal{R}^L_+$ and $p^t \in \mathcal{R}^L_{++}$ for all t.

If the consumer maximizes locally non-satiated utility u (defined on $\mathcal{R}^L_+$) subject to the budget constraint, then these observations indicate the following:

(1) her income in situation t is $p^t x^t$,

(2) $u(x^t) \geq u(x)$ for every x such that $p^t x \leq p^t x^t$,

(3) $u(x^t) > u(x)$ for every x such that $p^t x < p^t x^t$.

Note that local nonsatiation is crucial for (3). (2) and (3) imply that

$$\text{if } p^t x^s \leq p^t x^t, \quad \text{then } p^s x^t \geq p^s x^s \tag{23}$$

for all $s, t = 1, \ldots, T$.

Property (23) is called the **Generalized Weak Axiom of Revealed Preference**.

Utility function u **rationalizes** observations $\{(p^1, x^1), \ldots, (p^T, x^T)\}$ if, for every t, $u(x^t) \geq u(x)$ for every $x \in \mathcal{R}^L_+$ such that $p^t x \leq p^t x^t$.

We have just shown that GWARP necessarily holds for a set of observations rationalized by locally nonsatiated utility function. Is GWARP also a sufficient condition for rationalizability? The answer is **no.** To understand why, we take another look at what follows from utility maximization.

Let us define relations R and P between an observation x^t and a bundle $x \in \mathcal{R}_+^L$ as follows:

$$x^t R x, \quad \text{if} \ \ p^t x \leq p^t x^t, \tag{24}$$

$$x^t P x, \quad \text{if} \ \ p^t x < p^t x^t. \tag{25}$$

If $x^t R x$, we say that x^t is (directly) **weakly revealed preferred to** x.
If $x^t P x$, we say that x^t is (directly) **strictly revealed preferred to** x.

Again, if the consumer maximizes locally non-satiated utility u subject to the budget constraint, i.e., if the observations are rationalized by u, then

$x^t R x$ implies $u(x^t) \geq u(x)$, and

$x^t P x$ implies $u(x^t) > u(x)$.

Consequently, if $x^t R x^s$ then not $x^s P x^t$. This is the GWARP restated. But there is more. For every subset of observations $(p^{t_1}, x^{t_1}), \ldots, (p^{t_n}, x^{t_n})$,

$$\text{if } x^{t_1} R x^{t_2}, \ x^{t_2} R x^{t_3}, \ldots, \ x^{t_{n-1}} R x^{t_n}, \text{ then not } x^{t_n} P x^{t_1}. \tag{26}$$

Property (26) is called the Generalized Strong Axiom of Revealed Preference, or simply **Generalized Axiom of Revealed Preference,** GARP.

Theorem 13.1 (Afriat): *Observations* $(p^1, x^1), \ldots, (p^T, x^T)$ *satisfy GARP if and only if there exists a locally nonsatiated utility function* u *that rationalizes these observations.*

Proof: See Varian, Ch. 8, also Varian (1982). The utility function u is defined as follows: First, it is proved that the system of inequalities

$$u^t \leq u^s + \lambda^s p^s (x^t - x^s), \qquad \forall t, s.$$

has solution u^t, λ^t. Then, function u is defined by

$$u(x) = \min_t \{u^t + \lambda^t p^t (x^t - x)\}.$$

It holds $u(x^t) = u^t$. This function u is continuous, concave, and increasing.

GARP is a generalization of two standard "axioms." The **Weak Axiom of Revealed Preference** is

$$\text{if } x^t R x^s \text{ and } x^t \neq x^s, \text{ then not } x^s R x^t. \tag{27}$$

The **Strong Axiom of Revealed Preference** is

$$\text{if } x^{t_1} R x^{t_2},\ x^{t_2} R x^{t_3}, \ldots,\ x^{t_{n-1}} R x^{t_n} \text{ and } x^{t_1} \neq x^{t_n}, \text{ then not } x^{t_n} R x^{t_1}. \tag{28}$$

These axioms hold for observations **strictly rationalized** by utility function u, i.e, if $u(x^t) > u(x)$ for every $x \in \mathcal{R}^L_+, x \neq x^t$, such that $p^t x \leq p^t x^t$, for every t. SARP is strictly stronger (as long as $L > 2$) than WARP.

Remarks:

• Axioms of revealed preference can be used to derive algebraic comparative statics of Walrasian demand, see Varian, Ch.8.

• Axioms of revealed preference can also be used in the context of a demand function, instead of finite demand observations. MWG prove in Section 3.J that the Strong Axiom of Revealed Preference holds for a demand function satisfying budget equation if and only if there exists a rational preference relation $\succeq$ that rationalizes the demand function.

• There is a strong relationship between negative semi-definiteness of Slutsky matrix and axioms of revealed preference for a demand function. MWG prove in Section 2.F that if the WARP holds for a differentiable demand function d satisfying budget equation, then the Slutsky matrix associated with d is negative semi-definite.

PART III: Choice Under Uncertainty

14. Expected Utility with Objective Probabilities.

Let Z be a (finite) set of **outcomes**. We take $Z = \{z_1, \ldots, z_K\}$. A **lottery** on Z is a probability distribution on Z, that is, an assignment of probabilities $\{\pi_i\}_{i=1}^K$ to outcomes so that π_i is the probability of winning outcome z_i. Lottery with probabilities $\{\pi_i\}_{i=1}^K$ is denoted by L. Let $\mathcal{L}$ be the set of all lotteries on Z. Since probabilities add up to one and are positive, the set $\mathcal{L}$ can be identified with the unit simplex Δ^{K-1}.

For two lotteries $L, L' \in \mathcal{L}$ and a probability α such that $0 \leq \alpha \leq 1$, the **compound lottery** $\alpha L + (1-\alpha)L'$ is a lottery for which outcome z_i has probability $\alpha\pi_i + (1-\alpha)\pi_i'$, for every i. Often, compound lottery is denoted by $L\alpha L'$ instead of $\alpha L + (1-\alpha)L'$.

The decision maker has a complete and transitive preference relation $\succeq$ on the set of lotteries $\mathcal{L}$.

Preference relation $\succeq$ is said to have an **expected utility representation** if there exists function $v : Z \to \mathcal{R}$ (or simply K numbers $v_i = v(z_i)$ for all i) such that

$$L \succeq L' \quad \text{if and only if} \; \sum_{i=1}^K \pi_i v_i \geq \sum_{i=1}^K \pi_i' v_i.$$

Von Neumann–Morgenstern Axioms:

(i) *continuity,*

(ii) *independence:* for every $L, L', L'' \in \mathcal{L}$ and $0 \leq \alpha \leq 1$,

$$L \succeq L' \quad \text{if and only if } \alpha L + (1-\alpha)L'' \succeq \alpha L' + (1-\alpha)L''. \tag{29}$$

Theorem 14.1: *Preference relation $\succeq$ satisfies (**i**) and (**ii**) if and only if it has an expected utility representation.*

Proof: Outline.

Step 1: Versions of the independence axiom hold for indifference relation $\sim$ and strict preference $\succ$.

Step 2: There exist lotteries $\bar{L}$ and $\underline{L}$ such that $\bar{L} \succeq L \succeq \underline{L}$ for every $L \in \mathcal{L}$.

Step 3: If $L \succ L'$, then $L \succ \alpha L + (1-\alpha)L' \succ L'$ for every $0 < \alpha < 1$.

Step 4: $\beta\bar{L} + (1-\beta)\underline{L} \succ \alpha\bar{L} + (1-\alpha)\underline{L}$ if and only if $\beta > \alpha$.

Step 5: For every L there exists α_L such that $L \sim \alpha_L \bar{L} + (1-\alpha_L)\underline{L}$.

Step 6: Define $U(L) = \alpha_L$. This U represents $\succeq$ and is linear. Set $v_i = U(L^i)$, where L^i is the degenerate lottery that assigns probability 1 to z_i. Then $U(L) = \sum_i \pi_i v_i$ and is the desired expected utility representation.

Remarks:

• Function v of the expected utility representation is usually called **von Neumann-Morgenstern utility**, or sometimes, as in MWG, **Bernoulli utility.**

• vNM utility function v is unique up to an increasing affine transformation, that is, if preference relation $\succeq$ has expected utility representations with v and with $\tilde{v}$, then $\tilde{v}(z) = Av(z) + B$ for some A, B with $A > 0$.

• A difficulty for expected utility with objective probabilities – **Allais paradox.**

15. Expected Utility under Uncertainty

Uncertainty is described by a set $S = \{1, \ldots, S\}$ of *states of nature*. State-contingent consumption plan specifies consumption conditional on each state. We assume that there is a single commodity. Then consumption plan is a vector $c = (c_1, \ldots, c_S) \in \mathcal{R}^S$.

We consider preferences over state-contingent consumption plans described by a utility function $u : \mathcal{R}^S \to \mathcal{R}$. We assume that u *strictly increasing and continuous.*

We say that u has **state-separable representation** if there exist functions $v_s : \mathcal{R} \to \mathcal{R}$ for all s, such that

$$u(c_1, \ldots, c_S) \geq u(c'_1, \ldots, c'_S) \text{ iff } \sum_{s=1}^{S} v_s(c_s) \geq \sum_{s=1}^{S} v_s(c'_s) \tag{30}$$

for every c, c'.

We say that u has **expected utility representation** with respect to probabilities $\{\pi_s\}$ if there exists function $v : \mathcal{R} \to \mathcal{R}$ such that

$$u(c_1, \ldots, c_S) \geq u(c'_1, \ldots, c'_S) \text{ iff } \sum_{s=1}^{S} \pi_s v(c_s) \geq \sum_{s=1}^{S} \pi_s v(c'_s), \tag{31}$$

for every c, c'.

Utility function v in the expected utility representation is the von Neumann-Morgenstern (or Bernoulli) utility. Expected utility is written as $E[v(c)]$.

Axiomatization of State-Separable Utility

For $c \in \mathcal{R}^S$ and $y \in \mathcal{R}$, let $c_{-s}y$ denote the consumption plan c with consumption c_s in state s replaced by y, that is, the plan $(c_1, \ldots, c_{s-1}, y, c_{s+1}, \ldots, c_S)$.

The **independence axiom** (*sure-thing principle*):

$$u(c_{-s}y) \geq u(d_{-s}y) \qquad \text{iff } u(c_{-s}w) \geq u(d_{-s}w) \tag{32}$$

for all $c, d \in \mathcal{R}^S$ and $y, w \in \mathcal{R}$.

Theorem 15.1: *Assume that $S \geq 3$, and that utility function u is strictly increasing and continuous. Then u has a state-separable representation iff it obeys the independence axiom.*

Proof: see Debreu (1959), "Topological methods in cardinal utility theory".

Remark: This theorem does not hold for $S = 2$. With two states, the independence axiom is trivially satisfied by every strictly increasing utility function regardless of whether it has a state-separable representation or not.

For probabilities $\{\pi_s\}$ of states such that $\pi_s > 0$ for each s, let $E(c) = \sum_s \pi_s c_s$ be the expected value of $c = (c_1, \ldots, c_S)$ and let $\mathbf{E}(c)$ denote the deterministic consumption plan $(E(c), \ldots, E(c))$.

A consumer with utility function u is **risk averse** (with respect to $\{\pi_s\}$) if

$$u(c) \leq u(\mathbf{E}(c)) \tag{33}$$

for every c. Expected utility $E[v(\cdot)]$ is risk averse if and only if v is concave. (This will be proved later.)

Theorem 15.2: *Assume that $S \geq 3$, and that utility function u is strictly increasing and continuous. Then u satisfies the independence axiom and is risk averse with respect to probabilities $\{\pi_s\}$ with if and only if it has a concave expected utility representation with respect to $\{\pi_s\}$.*

Proof: The independence axiom implies that u has a state-separable representation $\sum_s v_s(c_s)$. Suppose that each function v_s is differentiable. [A proof without this extra assumption can be found in Werner (2005).]

For each $x \in \mathcal{R}$ consider the problem

$$\max_c \sum_s v_s(c_s) \tag{34}$$

subject to

$$E(c) = x.$$

By risk aversion, $c = (x, \dots, x)$ must be a solution to (34). FOCs for this solution are

$$v_s'(x) = \lambda \pi_s, \qquad s = 1, \dots S. \tag{35}$$

It follows form (35) that

$$v_s'(x) = \frac{\pi_s}{\pi_1} v_1'(x).$$

Since x was arbitrary, we obtain

$$v_s(x) = \frac{\pi_s}{\pi_1} v_1(x) + \Delta_s \tag{36}$$

for some Δ_s. Consequently, $\sum_s \pi_s v(c_s)$ with $v \equiv v_1$ represents the utility function u. Since u is risk averse, v is concave (again, this will be proved later).

Remarks: There are two alternative interpretations of the role of probabilities in the above theorem. One is that probabilities $\{\pi_s\}$ are objective probabilities. The other is that they are not objectively given, but the consumer displays risk aversion with respect to these π_s. That is, when faced with a choice between state-contingent plan c and deterministic plan $\mathbf{E}(c)$, where $\mathbf{E}(c)$ has been calculated using $\{\pi_s\}$, he or she prefers $\mathbf{E}(c)$. Then, the theorem says, probabilities $\{\pi_s\}$ are his "subjective" probabilities.

16. Risk Aversion and the Pratt's Theorem

A consumer with expected utility function is **risk averse** if

$$E[v(c)] \leq v(E(c)), \tag{37}$$

for every risky consumption plan c.

"Risky consumption plan" could mean either one of two things:

(I) State-contingent consumption $c = (c_1, \ldots, c_S) \in \mathcal{R}^S$ for $S \geq 2$ states with given probabilities $\{\pi_s\}$; then $E[v(c)] = \sum_s \pi_s v(c_s)$, and $E(c) = \sum_s \pi_s c_s$; or

(II) Real-valued random variable c with arbitrary distribution (described by cumulative distribution function F_c), that is, arbitrary lottery c with outcomes in $\mathcal{R}$ (as in MWG). Then $E[v(c)] = \int v(x) dF_c(x)$, and $E(c) = \int x dF_c(x)$.

The consumer is **strictly risk averse** if

$$E[v(c)] < v(E(c)) \tag{38}$$

for every risky consumption plan c such that $c \neq E(c)$. She is **risk neutral** if

$$E[v(c)] = v(E(c)) \tag{39}$$

for every risky consumption plan c.

Measures of Risk Aversion

The **risk compensation** for additional risky consumption plan z with $E(z) = 0$ at deterministic "initial" consumption $x \in \mathcal{R}$ is $\rho(x, z)$ that solves

$$E[v(x+z)] = v\big(x - \rho(x, z)\big). \tag{40}$$

If v is twice-differentiable and strictly increasing (so that $v'(x) > 0$ for every x), we also have:

– the Arrow-Pratt measure of **absolute risk-aversion**

$$A(x) \equiv -\frac{v''(x)}{v'(x)}, \tag{41}$$

– the Arrow-Pratt measure of **relative risk aversion**

$$R(x) \equiv -\frac{v''(x)}{v'(x)}x. \tag{42}$$

The Theorem of Pratt

Let v_1, v_2 be two C^2, strictly increasing vN-M. utility functions with ρ_1, ρ_2, and A_1 and A_2, respectively.

Theorem 16.1 (Pratt): *The following conditions are equivalent:*

(i) $A_1(x) \geq A_2(x)$ *for every* $x \in \mathcal{R}$*.*

(ii) $\rho_1(x, z) \geq \rho_2(x, z)$ *for every* $x \in \mathcal{R}$ *and every risky plan* z *with* $E(z) = 0$*.*

(iii) v_1 *is a concave transformation of* v_2*, i.e.* $v_1(x) = f(v_2(x))$ *for every* x*, for* f *concave and strictly increasing.*

Risk Aversion and Concavity

Let v be twice-differentiable and strictly increasing.

Corollary 16.2:

(i) *A consumer is risk averse iff his von Neumann-Morgenstern utility function v is concave.*

(ii) *A consumer is risk neutral iff his von Neumann-Morgenstern utility function v is linear.*

(iii) *A consumer is strictly risk averse iff his von Neumann-Morgenstern utility function v is strictly concave.*

Note: "*iff*" means "if and only if."

This corollary holds true even without the assumption of differentiability of v, see LeRoy and Werner(2001).

Decreasing, Constant and Increasing Risk Aversion

Corollary 16.3: Let v be C^2 and strictly increasing. Then

(i) *$\rho(x, z)$ is increasing in x for every z with $E(z) = 0$, iff $A(x)$ is increasing in x.*

(ii) *$\rho(x, z)$ is constant in x for every z with $E(z) = 0$, iff $A(x)$ is constant in x.*

(iii) *$\rho(x, z)$ is decreasing in x for every z with $E(z) = 0$, iff $A(x)$ is decreasing in x.*

Some Common Utility Functions

The functions most often used as von Neumann-Morgenstern utility functions in applied work and as examples are:

Linear utility:

$$v(x) = x$$

has zero absolute risk aversion, so the consumer is risk-neutral.

Negative Exponential Utility:

$$v(x) = -e^{-\alpha x},$$

where $\alpha > 0$, has constant absolute risk-aversion (CARA) equal to α.

Quadratic utility:

$$v(x) = -(\alpha - x)^2, \quad \text{for} \quad x < \alpha,$$

has absolute risk aversion equal to $1/(\alpha - x)$.

Logarithmic utility:

$$v(x) = ln(x + \alpha), \quad \text{for} \quad x > -\alpha.$$

If $\alpha = 0$, then relative risk-aversion is constant (CRRA).

Power utility:

$$v(x) = \frac{x^{1-\gamma}}{1-\gamma}, \quad \text{for} \quad x \geq 0,$$

where $\gamma \geq 0, \gamma \neq 1$, has constant relative risk-aversion equal (CRRA) to γ.

Linear Risk Tolerance

The **risk tolerance**:

$$T(x) \equiv \frac{1}{A(x)}.$$

The negative exponential utility function, the quadratic utility function, the logarithmic utility function, the power utility function — all have linear risk tolerance (LRT or HARA).

Proof of Pratt's Theorem 16.1:

(i) implies (iii): Define f by $f(t) = v_1(v_2^{-1}(t)$ for every t. The first derivative of f is

$$f'(t) = \frac{v_1'(v_2^{-1}(t))}{v_2'(v_2^{-1}(t))} \tag{43}$$

and is strictly positive since $v_i' > 0$ for $i = 1, 2$. The second derivative is

$$f''(t) = \frac{v_1''(x) - (v_2''(x)v_1'(x))/v_2'(x)}{[v_2'(x)]^2}, \tag{44}$$

where we used $x = v_2^{-1}(t)$. Equation (44) can be rewritten as

$$f''(t) = \big(A_2(x) - A_1(x)\big)\frac{v_1'(x)}{[v_2'(x)]^2}.$$

Thus $f''(t) \leq 0$ for every t, and hence f is concave.

(iii) implies (ii): By the definition of ρ_1 (see (40))

$$E[v_1(x+z)] = v_1\big(x - \rho_1(x,z)\big). \tag{45}$$

Since $v_1 = f(v_2)$ and f is concave, Jensen's inequality yields

$$E[v_1(x+z)] = E[f(v_2(x+z))] \leq f(E[v_2(x+z)]). \tag{46}$$

The right-hand side of (46) equals $f\big(v_2(x-\rho_2(x,z))\big)$ or $v_1(x-\rho_2(x,z))$. Using (45) and (46) we obtain

$$v_1\big(x - \rho_1(x,z)\big) \leq v_1\big(x - \rho_2(x,z)\big). \tag{47}$$

Since v_1 is strictly increasing, (47) implies that $\rho_1(x,z) \geq \rho_2(x,z)$.

(ii) implies (i): (... in class)

17. Stochastic Dominance and Risk

For a consumer whose preferences over state-contingent consumption plans in $\mathcal{R}^S$ have an expected utility representation, it is only the *probability distribution* of consumption that matters. That is, any two consumption plans that have the same probability distribution have the same expected utility. For instance, if there are two states with equal probabilities, then the expected utility of consumption plans $(1, 2)$ and $(2, 1)$ is the same.

Stochastic dominance is a ranking of random variables based on their distributions.

Let y and z be two random variables – for example, two state-contingent consumption plans. For simplicity, we assume that y and z take values in a bounded interval $[a, b]$. Let F_z and F_y be their **cumulative distribution functions**. That is,

$$F_z(t) = Prob(z \leq t)$$

for $t \in [a, b]$. We have

$$E(z) = \int_a^b t dF_z(t) \quad \text{and} \quad E[v(z)] = \int_a^b v(t) dF_z(t).$$

For the use later note that

$$E(z) = b - \int_a^b F_z(t) dt. \tag{48}$$

First-Order Stochastic Dominance

Definition 17.1: *z first-order stochastically dominates y if*

$$F_z(t) \leq F_y(t), \qquad \forall t \in [a, b]. \tag{49}$$

We have

Theorem 17.2: *z first-order stochastically dominates y if and only if*

$$E[v(z)] \geq E[v(y)]$$

for every nondecreasing continuous v.

That is, z FSD y if and only if every expected-utility maximizing agent with nondecreasing utility prefers z to y.

Example: Let y take values 1 and 3 with probabilities 1/2, and z take value 1 with probability 1/4, value 3 with probability 1/4, and value 4 with probability 1/2. Then z FSD y.

Definition 17.3: *z second-order stochastically dominates y if*

$$\int_a^w F_z(t)dt \leq \int_a^w F_y(t)dt, \quad \forall w \in [a, b]. \tag{50}$$

Inequality (50) for $w = b$ implies (using (48)) that $E(z) \geq E(y)$.

Thus, if z SSD y, then $E(z) \geq E(y)$. Also, if z FSD y, then z SSD y.

Theorem 17.4: *z second-order stochastically dominates y if and only if*

$$E[v(z)] \geq E[v(y)]$$

for every nondecreasing concave continuous v.

That is, z SSD y if and only if every agent with risk-averse nondecreasing expected utility prefers z to y.

If z SSD y and z and y have the same expectation $E(z) = E(y)$, then we say that y is **more risky**.

Proposition 17.5: If $E(z) = 0$, then $2z$ is more risky than z.

Proof: It suffices to prove that $E[v(z)] \geq E[v(2z)]$ for every nondecreasing concave v. Since $z = \frac{1}{2}(2z) + \frac{1}{2}(0)$, we have

$$\frac{1}{2}v(2z) + \frac{1}{2}v(0) \leq v(z). \tag{51}$$

Taking expectations on both sides of (51), we obtain

$$\frac{1}{2}E[v(2z)] + \frac{1}{2}v(0) \leq E[v(z)]. \tag{52}$$

Concavity of v and $E(z) = 0$ imply that

$$E[v(z)] \leq v(0). \tag{53}$$

Combining (52) with (53) we obtain $E[v(z)] \geq E[v(2z)]$.

Risk and Variance

For z and y with $E(z) = E(y)$, if y is more risky than z, then $\text{var}(y) \geq \text{var}(z)$. [This follows from $E[v(z)] \geq E[v(y)]$ applied to the quadratic utility $v(x) = -(\alpha - x)^2$.] The converse is not true!

Example: Let z take on the values $1, 3, 4, 6$ with equal probabilities, and let y take value 2 with probability 1/2 and values 3 and 7, each with probability 1/4. We have $E(z) = E(y) = 3.5$, and

$$\text{var}(y) = 4.25, \quad \text{var}(z) = 3.25.$$

Thus $\text{var}(y) > \text{var}(z)$. For the logarithmic utility $v(x) = \ln(x)$, we have

$$E[v(z)] = \frac{1}{4}\ln(72), \quad E[v(y)] = \frac{1}{4}\ln(84).$$

Thus, $E[v(z)] < E[v(y)]$. Since v is concave, it follows that y is not more risky than z. [In fact, neither y is more risky than z, nor z is more risky than y.]

Proof of Theorem 17.2 on First-Order Stochastic Dominance:

First, let $E[v(z)] \geq E[v(y)]$ for every nondecreasing continuous v. We want to show that $F_z(t) \leq F_y(t),\ \forall t \in [a, b]$. Suppose, by contradiction, that $F_z(t_0) > F_y(t_0)$ for some $t_0 \in [a, b]$. Define the following utility function

$$v(t) = \begin{cases} 0, & \text{if } \ t \leq t_0 \\ 1, & \text{if } \ t > t_0 \end{cases}$$

We have

$$E[v(z)] - E[v(y)] = F_y(t_0) - F_z(t_0) < 0.$$

Function v is nondecreasing, but it is not continuous. However, it can be approximated by a nondecreasing continuous function so that the expression $E[v(z)] - E[v(y)]$ remains strictly negative. This is a contradiction.

Second, let $F_z(t) \leq F_y(t), \quad \forall t \in [a, b]$. We want to show that $E[v(z)] \geq E[v(y)]$ for every nondecreasing continuous v. Suppose first that v is differentiable. We use intergation by parts:

$$E[v(z)] - E[v(y)] = \int_a^b v(t) dF_z(t) - \int_a^b v(t) dF_y(t) =$$

$$[v(b)F_z(b) - v(a)F_z(a)] - \int_a^b F_z(t) v'(t) dt - [v(b)F_y(b) - v(a)F_y(a)]$$

$$+ \int_a^b F_y(t) v'(t) dt = \int_a^b (F_y(t) - F_z(t)) v'(t) dt \geq 0.$$

The same argument holds without differentiability: see Tesfatsion (1976).

Proof of Theorem 17.4 on Second-Order Stochastic Dominance:

First, let $E[v(z)] \geq E[v(y)]$ for every nondecreasing continuous and concave v. We want to show that $\int_a^w F_z(t)dt \leq \int_a^w F_y(t)dt$ for all $w \in [a,b]$. Suppose, by contradiction, that $\int_a^{w_0} F_z(t)dt > \int_a^{w_0} F_y(t)dt$ for some w_0. Define the following utility function

$$v(t) = \begin{cases} t - w_0, & \text{if } \ t \leq w_0 \\ 0, & \text{if } \ t > w_0 \end{cases}$$

We have

$$E[v(z)] - E[v(y)] = \int_a^{w_0} (t - w_0)dF_z(t) - \int_a^{w_0} (t - w_0)dF_z(t) =$$

$$- \int_a^{w_0} F_z(t)dt + \int_a^{w_0} F_y(t)dt < 0,$$

where we used integration by parts. Function v is nondecreasing, continuous and concave. This is a contradiction.

Second, let $\int_a^w F_z(t)dt \leq \int_a^w F_y(t)dt$ for all $w \in [a,b]$. We want to show that $E[v(z)] \geq E[v(y)]$ for every nondecreasing continuous and concave v. Suppose first that v is twice-differentiable. We use the derivation from the proof of FSD and apply intergation by parts one more time:

$$E[v(z)] - E[v(y)] = \int_a^b (F_y(t) - F_z(t))v'(t)dt =$$

$$v'(b)[\int_a^b (F_y(t) - F_z(t))dt] - \int_a^b [\int_a^w (F_y(t) - F_z(t))dt] \ v''(w)dw \geq 0.$$

The same argument holds without differentiability: see Tesfatsion (1976).

Mathematical Appendix

I. Theorem of the Maximum

There are two sets $S \subset \mathbb{R}^n$ and $T \subset \mathbb{R}^m$. Further, there are a correspondence φ mapping S into the sets of subsets of T and a function $f : S \times T \to \mathbb{R}$. That is, $\varphi(x)$ is a subset of T for every $x \in S$, and $f(x,t)$ is a real number for every $x \in S$ and $t \in T$.

We are interested in the constrained maximization problem with f as the objective function and φ as the constraint. That is, given $x \in S$,

$$\max_t f(x,t) \tag{1}$$

$$\text{subject to} \qquad t \in \varphi(x).$$

We denote by $g(x)$ the maximized value of function f and by $\mu(x)$ the subset of vectors t in $\varphi(x)$ on which the maximum value is attained. Formally,

$$g(x) = \max_{t \in \varphi(x)} f(x,t) \quad \text{and} \quad \mu(x) = \{t \in \varphi(x) : f(x,t) = g(x)\}. \tag{2}$$

Interpretation: Think about an economic agent whose environment is described by a vector $x \in S$. The agent's set of actions is T, but when the environment is x, she is restricted to choose her action only from the subset $\varphi(x)$. Her utility of action t is $f(x,t)$, when the environment is x. Her objective is to choose an action in $\varphi(x)$ to maximize her utility.

We shall assume that the set T is **compact.**

Correspondence φ is said to be **continuous** if it is lower hemi-continuous and upper hemi-continuous. These are defined as follows:

(LHC) φ is **lower hemi-continuous** at x if for every sequence $\{x_n\}$ in S converging to x and every $t \in \varphi(x)$, there exists a sequence $\{t_n\}$ in T such that $t_n \in \varphi(x_n)$ and $\{t_n\}$ converges to t.

(UHC) φ is **upper hemi-continuous** at x if for every sequence $\{x_n\}$ in S converging to x and every sequence $\{t_n\}$ in T converging to t, with $t_n \in \varphi(x_n)$, it holds that $t \in \varphi(x)$.

Our definition of UHC is the closed graph property. MasColell, Whinston and Green give definitions of LHC and UHC in Appendix M.H, pg. 949-951. Their definition of upper hemi-continuity is different, but if the range of φ (i.e., the set T) is compact as assumed, then their definition is equivalent to the above one. Note that upper hemi-continuous correspondence φ must have compact values $\varphi(x)$.

Theorem I.1: *Suppose that the set T is compact. If correspondence φ is continuous on S and function f is continuous on $S \times T$, then*

(i) g is continuous on S, and

(ii) μ is an upper hemi-continuous correspondence on S.

Proof: (i) Let $\{x_n\}$ be a sequence of vectors in S converging to x. We have to show that $\lim_n g(x_n) = g(x)$. Since $\varphi(x_n)$ is a compact set for every n, there exist $t_n \in \varphi(x_n)$ such that $g(x_n) = f(x_n, t_n)$. Since the set T is compact, sequence $\{t_n\}$ must have a convergent subsequence with a limit $\bar{t} \in T$. We switch to that subsequence of $\{t_n\}$, but we retain the same notation; i.e., we keep $\{t_n\}$ and assume that it converges to $\bar{t}$. Upper hemi-continuity of φ implies that $\bar{t} \in \varphi(x)$. By continuity of f, we have $\lim_n f(x_n, t_n) = f(x, \bar{t})$. Since $f(x, \bar{t}) \leq g(x)$, it follows that

$$\lim_n g(x_n) \leq g(x).$$

To prove the oposite inequality, we note that $g(x) = f(x,t)$ for some $t \in \varphi(x)$ since $\varphi(x)$ is a compact set. Lower hemi-continuity of φ at x implies that there is sequence $\{\tilde{t}_n\}$ converging to t such that $\tilde{t}_n \in \varphi(x_n)$ for every n. We have $f(x_n, \tilde{t}_n) \leq g(x_n)$. Using continuity of f, we obtain $\lim_n f(x_n, \tilde{t}_n) = f(x,t)$. Consequently

$$\lim_n g(x_n) \geq g(x).$$

This concludes the proof of (i)

(ii) Consider two sequences: $\{x_n\}$ in S converging to x, and $\{t_n\}$ in T converging to t such that $t_n \in \mu(x_n)$. We have to show that $t \in \mu(x)$.

We first observe that upper hemi-continuity of φ implies that $t \in \varphi(x)$. Next, consider arbitrary $\tilde{t} \in \varphi(x)$. Lower hemi-continuity of φ at x implies that there is a sequence $\{\tilde{t}_n\}$ converging to $\tilde{t}$ such that $\tilde{t}_n \in \varphi(x_n)$ for every n. Clearly then $f(x_n, t_n) \geq f(x_n, \tilde{t}_n)$. Passing to the limit with n and using continuity of f, we obtain $f(x,t) \geq f(x,\tilde{t})$. Since $\tilde{t}$ was arbitrary in $\varphi(x)$, this implies that $t \in \mu(x)$. This concludes the proof of (ii).

Remarks:

- The assumption that set T is compact can be dropped. Then the MWG definition of upper hemi-continuity has to be used. Note that this definition requires that correspondence φ is compact-valued.

- One application of the Theorem of the Maximum I.1 is in producer theory. We set S as the set of price vectors, T as the production set, i.e., $T = Y$, function f as $f(p,y) = py$, and correspondence φ as $\varphi(p) = Y$. Assuming that Y is compact, Theorem I.1 implies continuity of the profit function (Theorem 4.1 (i)) and upper hemi-continuity of the supply correspondence (Proposition 4.3 (iii)).

II. Kuhn-Tucker Theorems

II.1 Constrained Maximization: Necessary Conditions.

Function $F : \mathbb{R}^n_+ \to \mathbb{R}$ is the objective function; functions $g^j : \mathbb{R}^n_+ \to \mathbb{R}$, for $j = 1, \ldots, k$, are constraint functions. Assume that F and g^j are differentiable, with partial derivatives $\frac{\partial F}{\partial x_i}$ and $\frac{\partial g^j}{\partial x_i}$ denoted by $\partial_i F$ and $\partial_i g^j$, respectively.

The constrained maximization problem (with nonnegativity constraints) is

$$\max_x F(x) \tag{3}$$

$$\begin{aligned} \text{subject to} \qquad & g^1(x) \geq 0, \\ & \ldots\ldots, \\ & g^k(x) \geq 0, \\ & x_1 \geq 0, \ldots, x_n \geq 0. \end{aligned}$$

We write the Lagrangian as

$$\mathcal{L}(\lambda^1, \ldots, \lambda^k, x) = F(x) + \sum_{j=1}^{k} \lambda_j g^j(x),$$

where $\lambda_j \geq 0$, for $j = 1, \ldots, k$, are the Lagrange multipliers. We use λ to denote the k-vector of multipliers.

Kuhn-Tucker conditions for $x^* \geq 0$ and $\lambda^* \geq 0$ are:

$$\text{for all } i = 1, \ldots, n \text{ and } j = 1, \ldots, k,$$

$$\partial_i F(x^*) + \sum_{j=1}^{k} \lambda_j^* \partial_i g^j(x^*) \leq 0, \quad \text{and if } x_i^* > 0, \text{ then `` } = 0\text{''}, \tag{4a}$$

$$g^j(x^*) \geq 0, \quad \text{and if } \lambda_j^* > 0, \text{ then `` } = 0\text{''}. \tag{4b}$$

Where do these conditions come from? Think about *maximizing* Lagrangian $\mathcal{L}(\lambda, x)$ with respect to x and *minimizing* it with respect to λ, unconstrained, except for $x \geq 0$ and $\lambda \geq 0$. This is the saddle-point. K-T conditions (4) are FOCs for such max-min (or saddle-point) problem.

Theorem (Kuhn-Tucker): *If $x^* \geq 0$ is a solution to the constrained maximization problem, and the Constraint Qualification Condition holds, then x^* and some $\lambda^* \geq 0$ satisfy K-T conditions (4).*

Constraint Qualification Condition:

(i) Kuhn-Tucker original – don't touch it.

(ii) g^j concave for all j, and **Slater's condition**, that is, there is some $x^0 \geq 0$ with $g^j(x^0) > 0$ for all j.

(iii) rank condition (see Takayama 1.D.4, or Varian, ch 27),

(iv) g^j linear for all j, (Arrow-Hurwicz-Uzawa, see Takayama 1.D.4)

II.2 Sufficiency of Kuhn-Tucker Conditions.

The most standard theorem is:

Theorem S1: *Suppose that F and $g^1, \ldots, g^k$ are all concave functions. If $x^* \geq 0$ and $\lambda^* \geq 0$ satisfy K-T conditions (4), then x^* is a solution to the constrained maximization problem.*

A better theorem is due to Arrow and Enthoven (1961).

Theorem S2: *Suppose that F and $g^1, \ldots, g^k$ are all quasi-concave functions and some "mild" condition holds. If $x^* \geq 0$ and $\lambda^* \geq 0$ satisfy K-T conditions (4), then x^* is a solution to the constrained maximization problem.*

The extra ("mild") condition is not needed if F is concave (and $g^1, \ldots, g^k$ are quasi-concave). See Takayama 1.E for three versions of the condition.

Quasi-concavity (and therefore also concavity) of functions g^j implies that the constraint set, i.e. the set of $x \geq 0$ satisfying $g^1(x) \geq 0, \ldots, g^k(x) \geq 0$, is convex.

II.3 Constrained Minimization

The constrained minimization problem (with nonnegativity constraints) is

$$\min_x F(x) \tag{5}$$

$$\text{subject to} \qquad g^1(x) \le 0, \ \ldots, \quad g^k(x) \le 0,$$

$$x_1 \ge 0, \ldots, x_n \ge 0.$$

The Lagrangian is

$$\mathcal{L}(\lambda, x) = F(x) + \sum_{j=1}^{k} \lambda_j g^j(x).$$

Kuhn-Tucker conditions for $x^* \ge 0$ and $\lambda^* \ge 0$ are,

for all $i = 1, \ldots, n$ and $j = 1, \ldots, k$,

$$\partial_i F(x^*) + \sum_{j=1}^{k} \lambda_j^* \partial_i g^j(x^*) \ge 0, \quad \text{and if } x_i^* > 0, \text{ then "} = 0\text{"}, \tag{6a}$$

$$g^j(x^*) \le 0, \quad \text{and if } \lambda_j^* > 0, \text{ then "} = 0\text{"}. \tag{6b}$$

The corresponding saddle-point problem is to *minimize* Lagrangian $\mathcal{L}(\lambda, x)$ with respect to x and *maximize* it with respect to λ for $x \ge 0$ and $\lambda \ge 0$.

The Kuhn-Tucker Theorem holds with no change for the constrained minimization problem. However, in constraint qualification conditions concavity of functions g^j, if present, has to be replaced by their convexity. This guarantees convexity of the constraint set described here by inequalities $g^j(x) \le 0$.

Theorems S1 and S2 continue to hold with concavity (quasi-concavity) of functions F and g^j replaced by their convexity (quasi-convexity, respectively).

II.4 Remarks:

• **Applications** of K-T theorems in microeconomics:

(i) Consumer theory: utility maximization subject to budget constraint, and expenditure minimization.

(ii) Welfare economics: Characterization of Pareto optimal allocations as solutions to maximization of a welfare function subject to resource constraints, or maximization of one agent's utility subject to constraints on other agents' utilities and resource constraints.

(iii) Producer theory: cost minimization.

• There are versions of K-T theorems for maximization and minimization with mixed constraints, i.e., when some constraints are of the equality form, $g^j(x) = 0$. See Sundaram [2], Section 6.4, or Adam Ślawski's notes.

• K-T theorems hold for *local* maxima (minima) as well.

References: [**1**] Takayama (1995), 1.D and 1.E. [**2**] Sundaram (1999), Chapter 6. [**3**] Varian, ch 27. [**4**] MasColell et al. *Warning:* Takayama [1] and Sundaram [2] do not explicitly write nonnegativity constraints $x \geq 0$. Varian [3] writes constraints in the maximization problem as $g^j(x) \leq 0$.

II.5 Example: Consider the following constrained maximization problem:

$$\begin{aligned} \text{maximize} \quad & \ln(x_1+1)+\ln(x_2+1) \\ \text{subject to} \quad & p_1x_1+p_2x_2 \le m \\ & x_1 \ge 0, \quad x_2 \ge 0, \end{aligned}$$

where $p_1 > 0$, $p_2 > 0$ and $m > 0$.

In order to derive the solution (as a function of parameters p_1, p_2 and m) we write the Kuhn-Tucker first-order conditions (4) as

$$(1) \qquad \frac{1}{x_1^*+1} - \lambda^* p_1 \le 0, \quad \text{and if} \quad x_1^* > 0, \text{ then "} = 0\text{".}$$

$$(2) \qquad \frac{1}{x_2^*+1} - \lambda^* p_2 \le 0, \quad \text{and if} \quad x_2^* > 0, \text{ then "} = 0\text{".}$$

$$(3) \qquad p_1x_1^* + p_2x_2^* \le m, \quad \text{and if } \lambda^* > 0, \text{ then "} = 0\text{".}$$

with $x^* \ge 0$ and $\lambda^* \ge 0$.

Note that (3) holds with equality since it follows from (1) that $\lambda^* > 0$.

We solve inequalities (1-3) by considering cases:

Case 1. $x_1^* > 0$, $x_2^* > 0$.

Then (1) and (2) hold with equalities. Solving (1), (2) and (3) we find $x_1^* = \dfrac{m+p_2-p_1}{2p_1}$ and $x_2^* = \dfrac{m+p_1-p_2}{2p_2}$ and $\lambda^* = \dfrac{2}{m+p_1+p_2}$. For x_1^* and x_2^* to be strictly positive, it has to be that $m+p_2 > p_1$ and $m+p_1 > p_2$. Thus Case 1 applies with x_1^* and x_2^* as listed above if $m+p_2 > p_1$ and $m+p_1 > p_2$.

Case 2. $x_1^* > 0$, $x_2^* = 0$.

(3) implies that $x_1^* = \dfrac{m}{p_1}$. Since (1) holds with equality, we solve it for $\lambda^* = \dfrac{1}{m+p_1}$. Next we need to verify inequality (2). It states

$$1 - \frac{p_2}{m+p_1} \leq 0,$$

and it holds if $p_2 \geq m + p_1$. Thus Case 2 applies (with $x_1^* = \dfrac{m}{p_1}$, $x_2^* = 0$) if $p_2 \geq m + p_1$.

Case 3. $x_1^* = 0$, $x_2^* > 0$.

This case is very similar to Case 2. From (3) and (2) we obtain $x_1^* = \dfrac{m}{p_2}$, $\lambda^* = \dfrac{1}{m+p_2}$. Verifying inequality (1), we obtain $p_1 \geq m + p_2$. Thus Case 3 applies (with $x_1^* = 0$, $x_2^* = \dfrac{m}{p_2}$) if $p_1 \geq m + p_2$.

The case $x_1^* = x_2^* = 0$ cannot hold since it violates equation (3). This concludes our solution to the K-T conditions.

Since utility function is concave and the constraint function is concave (in fact, it is linear) $K - T$ conditions are sufficient (Theorem S1). Hence, the solution to K-T conditions is a constrained maximizer. Further, since the Slater's condition holds, every constrained maximizer has to satisfy $K - T$ conditions.

III. Monotonicity of a Vector-Valued Function.

Let D be an open convex subset of $\mathbb{R}^n$, and let $F : D \to \mathbb{R}^n$.

Proposition III.1: *Suppose that F is continuously differentiable. Then the following two conditions are equivalent:*

(i) $[F(x') - F(x)][x' - x] \geq 0$ *for every* $x, x' \in D$,

(ii) *the matrix* $DF(x)$ *is positive semi-definite for every* $x \in D$.

Proof: Consider arbitrary $x, x' \in D$, and denote $x' - x$ by z. Further, define

$$x(\lambda) = \lambda x' + (1 - \lambda)x, \qquad \text{for } \lambda \in [0, 1]$$

It holds, $x(0) = x$, $x(1) = x'$, and $x(\lambda) = x + \lambda z$. Next, define function $g : [0, 1] \to \mathbb{R}$ by

$$g(\lambda) = z[F(x(\lambda)) - F(x)].$$

Note that $g(0) = 0$, $g(1) = [x' - x][F(x') - F(x)]$, and $g'(\lambda) = zDF(x(\lambda))z$. Suppose that (i) holds. Since $g(\lambda) = \frac{1}{\lambda}[x(\lambda) - x][F(x(\lambda)) - F(x)]$ for $\lambda > 0$, it follows that $g(\lambda) \geq 0$. Therefore, g has a minimum at $\lambda = 0$. This implies $g'(0) \geq 0$, which is $zDF(x)z \geq 0$. Since z was arbitrary, we obtain (ii).

Conversely, suppose that (ii) holds. Then $g'(\lambda) \geq 0$ for every $\lambda \in [0, 1]$. So function g is increasing and hence $g(1) \geq g(0) = 0$. This implies (i). QED