PRODUCTION

APEC 8002, Fall 2016 **Instructor: Terry Hurley**

INTRODUCTION (MWG 5.A-B; Varian 1; Chambers 1; Cornes 5.1,5.8)

Having completed your exploration of consumer behavior and demand, it is time to consider producer behavior and supply. To start, we will discuss the idea of combining commodities to produce new commodities, which is basically the notion of production functions. We will then consider how producers determine how to produce and how much to produce. We will look at these questions from several perspectives: cost minimization, revenue maximization, and profit maximization. With these perspectives, we will develop the concepts of cost functions, conditional input demand, revenue functions, conditional supply, profit functions, supply (unconditional), and input demand (unconditional). We will discuss the properties of these objects and how they are related. Our ultimate goal is the careful derivation of the supply side of a market economy and identification of useful relationships that can help guide our empirical work.

NOTATION

M:	Number of commodities that are outputs.
N:	Number of commodities that are inputs.
L:	Number of commodities including those that must be produced from other commodities (e.g., cloth and shirts) as well as primary factors (e.g., land and labor) such that $L = M + N$.
$\mathbf{y} \in \mathbb{R}^L$:	Production vector.
y_l:	Particular element of a production vector ($y_l > 0$ for a net output, $y_l < 0$ for a net input, and $y_l = 0$ when whatever is produced is used up or the commodity is not used at all).
$\mathbf{q} \in \mathbb{R}_+^M$:	Vector of outputs.
$q_m \geq 0$:	Specific output.
$\mathbf{z} \in \mathbb{R}_+^N$:	Vector of inputs (also called factors).
$z_n \geq 0$:	Specific input.
$\mathbf{p} \in \mathbb{R}_{++}^L$:	Vector of prices.

$\mathbf{p} \in \mathbb{R}_{++}^{M}$: Vector of output prices.

$p_l > 0$: Specific price/output price.

$\mathbf{r} \in \mathbb{R}_{++}^{N}$: Vector of input prices.

$r_n > 0$: Specific input price.

Note that I will denote vectors and sets using bold and scalars using italics. The symbol $\mathbb{R}$ tells us that we are using real numbers. The superscript on this symbol tells us we are looking at a vector of real numbers with the number of elements in the vector equal to the superscript. The subscript tells us if we are restricting the elements in the vector: + means all the elements are non-negative, ++ means they are all positive, – means they are all non-positive, and – – means they are all negative.

DESCRIPTIONS OF PRODUCTIVE POSSIBILITIES

Most of you are likely familiar with production in the context of a single output and many inputs where there is a clear delineation between outputs and inputs. We will review this special case in the process of developing more general descriptions. While reviewing the classic single output and many inputs case, it is instructive to pay close attention to the similarities and differences between this theory and the theory of consumer behavior.

We begin with the notion of the *Production Possibilities Set*. This set represents all feasible combinations of outputs and inputs. We will label it **PPS**. In general, **PPS** $\subset \mathbb{R}^{L}$. For our single output and many input case, **PPS** $\subset \mathbb{R}_{+} \times \mathbb{R}_{-}^{N}$ (all vectors with a non-negative first element and N non-positive elements). To be clear about how we are doing things, $\mathbf{y} = (\mathbf{q}, -\mathbf{z})$ where net outputs are measured as positive numbers and net inputs are measured as negative numbers (remember we have restricted $\mathbf{z}$ to be a vector of non-negative inputs, so to make them a negative net output, we need to multiply by -1). However, it is also important to realize, that what is an output and what is an input can change based on prices in the most general representations we will discuss. If $\mathbf{y}$ is a feasible option, $\mathbf{y} \in$ **PPS**.

The most talked about factor determining the **PPS** is technology, but it is also reasonable to think about **PPS** being constrained by institutional factors (e.g., unions may negotiate minimum staffing requirements). Another common distinction is time. Inputs like labor can often be immediately reallocated, while capital inputs like buildings and machinery can take longer to reallocate. This leads to a distinction between *Long-Run* and *Short-Run* production (or more generally *Unrestricted* and *Restricted* production). In the short-run, some commodities cannot be reallocated, which further restricts a producer's production possibilities. In the long-run, all commodities can be reallocated, which gives a producer maximum flexibility. Of course the black and white distinction between long-run and short-run production is really grey and typically determined by the application. For now, it is enough to realize that in the short-run we can define $\mathbf{y} = (\mathbf{y}^v, \mathbf{y}^f)$ where $\mathbf{y}^v$ is a vector of variable commodities that can be reallocated and $\mathbf{y}^f$ is a vector of fixed commodities that cannot be reallocated. The production possibilities set can then be written as conditional on fixed commodities such that $\mathbf{y}^v \in \mathbf{PPS}(\mathbf{y}^f)$ in the short-run and $(\mathbf{y}^v, \mathbf{y}^f) \in \mathbf{PPS}$ in the long-run. Below, we will simply refer to $\mathbf{y}$ and the **PPS**, though it is important to keep in mind that everything we say can also apply to $\mathbf{y}^v$ and $\mathbf{PPS}(\mathbf{y}^f)$. With that said, while some assumptions we introduce below make sense in the long-run, they may not make sense in the short-run.

Unrestricted and *Restricted* production are generalizations of long-run and short-run production often seen in the literature. *Unrestricted* is directly analogous to long-run production where a producer has maximum flexibility to choose all inputs and outputs. *Restricted* production assumes that there is some sort of constraint that limits a producer's options such that if $\mathbf{PPS}^R$ is the restricted production possibilities set then $\mathbf{PPS}^R \subset \mathbf{PPS}$ (the restricted production possibility set is a subset of the unrestricted one).

With the **PPS** defined, we can define a couple of other concepts, some of which should be familiar. Note that in the definitions below and throughout this class $\mathbf{x} \geq \mathbf{x}'$ where $\mathbf{x}$ and $\mathbf{x}'$ are vectors means that each element of $\mathbf{x}$ is greater than or equal to the corresponding element of $\mathbf{x}'$. When we write $\mathbf{x} \geq \mathbf{x}'$ and $\mathbf{x} \neq \mathbf{x}'$, we mean that $\mathbf{x}$ has at least one element greater than $\mathbf{x}'$ and no smaller elements. Also note that when we write a statement like $\mathbf{x} \geq \mathbf{0}$, $\mathbf{0}$ represents a vector of $\mathbf{0}$s with the number of elements equal to the number of elements in $\mathbf{x}$.

Input Requirements Set: All combinations of inputs capable of producing a particular combination of output: **IRS**(**q**) = {**z** ∈ $\mathbb{R}_+^N$: (**q**, -**z**) ∈ **PPS**}.

Isoquant: All combinations of the least amount of inputs required to produce a particular combination of output: **ISQ**(**q**) = {**z** ∈ $\mathbb{R}_+^N$: **z** ∈ **IRS**(**q**) and **z**' ∉ **IRS**(**q**) for all **z** ≥ **z**' and **z**' ≠ **z**}.

The input requirement set is directly analogous to the upper contour set defined for the consumer's problem, while an isoquant is directly analogous to the indifference contour set. Note that these concepts assume that outputs and inputs are clearly delineated.

Feasible Output Set: All combinations of outputs that can be produced from a given combination of inputs: **FOS**(**z**) = {**q** ∈ $\mathbb{R}_+^M$: (**q**, -**z**) ∈ **PPS**}.

Production Possibilities Frontier: All combinations of the most possible output obtainable from a particular input combination: **PPF**(**z**) = {**q** ∈ $\mathbb{R}_+^M$: **q** ∈ **FOS**(**z**) and **q**' ∉ **FOS**(**z**) for all **q**' ≥ **q** and **q**' ≠ **q**}.

The feasible output set is directly analogous to the input requirement set, while the production possibilities frontier is directly analogous to the isoquant. The only difference is that in one case we fix outputs, while in the other we fix inputs. Again, a clear delineation of outputs and inputs is assumed.

Technically Efficient: A production vector is technically efficient if there is no way to produce more output with the same amount of inputs and no way to use fewer inputs to produce the same amount of output:

- **y** ∈ **PPS** such that **y**' ∉ **PPS** for all **y**' ≥ **y** and **y**' ≠ **y** in general or
- (**q**, -**z**) ∈ **PPF**(**z**) and (**q**, -**z**) ∈ **ISQ**(**q**) when outputs and inputs are clearly delineated.

Transformation Function: A real valued function T: $\mathbb{R}^L \rightarrow \mathbb{R}$ where $T(\mathbf{y}) = 0$ if and only if $\mathbf{y}$ is technically efficient.

The transformation function provides as general of a description of efficient production as we will need in this course. For example, it is perfectly suited for a firm that uses cotton to produce cloth, sells some of the cloth to other firms, and makes shirts and pants to sell to stores with the remaining cloth. However, there are many applications where all this generality may be more than what is needed and may even cloud the issue. Indeed, many researchers have made plenty of progress with the simpler notion of the production function $q = f(\mathbf{z})$, which relates many inputs to the efficient production of a single output and implies $T(\mathbf{y}) = T(q, -\mathbf{z}) = f(\mathbf{z}) - q$.

When researches have needed to tackle questions regarding many outputs and many inputs where there is a clear distinction between outputs and inputs, the concept of a distance function has proven useful. While distance functions come in many flavors and levels of generality, we will focus on the input and output distance functions first introduced by Shephard. With an understanding of these two types of distance functions, you should be prepared to tackle more recent generalizations.

To build intuition for the input distance function, consider the single input, single output **PPS** illustrated in Figure 1. Consider point a, which corresponds to the output q^0 and input z^2. This choice of output and input is feasible since it falls in the **PPS**. However, it is not technically efficient because by proportionally reducing z^2 until it just equals z^1, we could still produce q^0. Alternatively, consider point b which corresponds to the output q^0 and input z^0. This choice is not feasible because it falls outside of the **PPS**. However, if we proportionally scale z^0 up to z^1, we would be able to produce q^0 efficiently. With this intuition, consider the function

P1 $$D_I(\mathbf{q}, \mathbf{z}) = \max_\delta \left\{\delta > 0 : \left(\mathbf{q}, -\frac{\mathbf{z}}{\delta}\right) \in \mathbf{PPS}\right\} = \max_\delta \left\{\delta > 0 \middle| \frac{\mathbf{z}}{\delta} \in \mathbf{IRS}(\mathbf{q})\right\}.$$

$D_I(\mathbf{q}, \mathbf{z})$ is referred to as the *input distance function*. Intuitively, it measures how much we need scale inputs proportionally up or down to efficiently produce $\mathbf{q}$. Applying this function to points

a, b, and c yields $D_I(q^0, z^2) = \frac{z^2}{z^1} > 1$, $D_I(q^0, z^0) = \frac{z^0}{z^1} < 1$, and $D_I(q^0, z^1) = \frac{z^1}{z^1} = 1$. It is worth noting that the input distance function yielded a value greater than 1 for our technically inefficient input choice, less than 1 for our infeasible input choice, and equal to 1 for our technically efficient input choice. Must this relationship always hold? It wouldn't, if **PPS** was represented by the unshaded region above our curve rather than the shaded region below it, but that would imply more output could be produced with fewer inputs, which seems nonsensical. Below, we will see that this relationship will indeed hold for typical assumptions used to characterize the **PPS** such that we can write the transformation function as $T(\mathbf{q}, -\mathbf{z}) = 1 - D_I(\mathbf{q}, \mathbf{z})$ (while we could technically also write $T(\mathbf{q}, -\mathbf{z}) = 1 - D_I(\mathbf{q}, \mathbf{z}) - 1$, there are reasons in terms of economic interpretation why it is better not to). This input distance function generalizes nicely to vectors of outputs and inputs, though if we want to be truly rigorous, we should replace the max in equation P1 with sup (the supremum) because it may not be possible to actually reach the maximum even under common assumptions.

To avoid getting bogged down in mathematical minutiae that goes beyond the prerequisites of this course, we will focus on distance functions that have well defined extremum, which will permit us to use max and min instead of sup and inf. If you are interested in a more detailed treatment, you can always work through Fare and Primont (1995) in your spare time.

To make the notion of the input distance function more concrete, we can derive these functions for a simple two output and two input production possibilities set:

$$\mathbf{PPS} = \{(\mathbf{q}, -\mathbf{z}) \in \mathbb{R}_+^2 \times \mathbb{R}_-^2 : \sqrt{z_1 z_2} \geq {q_1}^2 + {q_2}^2\}.$$

The input distance function is then defined by

$$D_I(\mathbf{q}, \mathbf{z}) = \max_\delta \left\{\delta > 0 : \sqrt{\left(\frac{z_1}{\delta}\right)\left(\frac{z_2}{\delta}\right)} \geq {q_1}^2 + {q_2}^2\right\}.$$

Note that $\sqrt{\left(\frac{z_1}{\delta}\right)\left(\frac{z_2}{\delta}\right)} \geq q_1{}^2 + q_2{}^2$ implies $\sqrt{z_1 z_2} \geq \delta(q_1{}^2 + q_2{}^2)$ for **q** and **z** to be feasible.

Now if $q_1 > 0$ or $q_2 > 0$, we can also write $\frac{\sqrt{z_1 z_2}}{(q_1{}^2+q_2{}^2)} \geq \delta$, which makes it immediately clear that the maximum value δ can take and still be feasible is $\frac{\sqrt{z_1 z_2}}{q_1{}^2+q_2{}^2}$. However, if $q_1 = 0$ and $q_2 = 0$, then we have $\sqrt{z_1 z_2} \geq 0$, which is true for any $\delta > 0$. In this instance, a maximum does not technically exist, which explains why a more rigorous treatment of input distance functions uses the supremum. So with the caveat of having $q_1 > 0$ or $q_2 > 0$, $D_I(\mathbf{q}, \mathbf{z}) = \frac{\sqrt{z_1 z_2}}{q_1{}^2+q_2{}^2}$.

There is a useful property of an input distance function that follows immediately from its definition:

The input distance function is homogeneous of degree 1 in inputs.

To see why this is true, if the input distance function is homogeneous of degree 1 in inputs, $D_I(\mathbf{q}, \mu\mathbf{z}) = \mu D_I(\mathbf{q}, \mathbf{z})$ for any $\mu > 0$. By definition, $D_I(\mathbf{q}, \mu\mathbf{z}) = \max_\delta \left\{\delta > 0: \left(\mathbf{q}, -\frac{\mu\mathbf{z}}{\delta}\right) \in \mathbf{PPS}\right\}$ where $\max_\delta \left\{\delta > 0: \left(\mathbf{q}, -\frac{\mu\mathbf{z}}{\delta}\right) \in \mathbf{PPS}\right\}$ can also be written as $\max_{\frac{\mu\delta}{\mu}} \left\{\frac{\mu\delta}{\mu} > 0: \left(\mathbf{q}, -\frac{\mu\mathbf{z}}{\delta}\right) \in \mathbf{PPS}\right\}$ or $\mu \max_{\frac{\delta}{\mu}} \left\{\frac{\delta}{\mu} > 0: \left(\mathbf{q}, -\frac{\mathbf{z}}{\frac{\delta}{\mu}}\right) \in \mathbf{PPS}\right\} = \mu D_I(\mathbf{q}, \mathbf{z})$.

ASIDE: Homogeneity properties of functions are widely exploited in economics, so it is worth defining the concept now that we have introduced a special case. Consider a function $f(\mathbf{x}, \mathbf{y})$ where **x** and **y** are real vectors. If this function is homogenous of degree k in **x**, then $f(\alpha\mathbf{x}, \mathbf{y}) = \alpha^k f(\mathbf{x}, \mathbf{y})$ for all $\alpha > 0$. Two other important properties are also often employed. If $f(\mathbf{x}, \mathbf{y})$ is differentiable and homogenous of degree k in **x**:

(i) $\sum_{i=1}^{N} \frac{\partial f(\mathbf{x},\mathbf{y})}{\partial x_i} x_i = kf(\mathbf{x}, \mathbf{y})$ where N is the number of elements in **x** (indeed, this is an if and only if proposition), and

(ii) $\frac{\partial f(\mathbf{x},\mathbf{y})}{\partial x_i}$ is homogeneous of degree $k - 1$ in **x**.

The *output distance function* is defined similarly to the *input distance function*:

P2 $$D_O(\mathbf{q},\mathbf{z}) = \min_\delta \left\{\delta > 0 : \left(\frac{\mathbf{q}}{\delta}, -\mathbf{z}\right) \in \mathbf{PPS}\right\} = \min_\delta \left\{\delta > 0 \middle| \frac{\mathbf{q}}{\delta} \in \mathbf{FOS}(\mathbf{z})\right\}.$$

A rigorous treatment would replace the min in equation P2 with inf (the infimum) because it may not be possible to reach the minimum under common assumptions. Intuitively, the output distance function measures how much we need to proportionally scale the output vector **q** up or down to achieve efficient production with inputs **z**.

COMMON ASSUMPTIONS OF PRODUCTIVE POSSIBILITIES

The description of the **PPS** provided above is a useful start, but it is too general to get us very far in terms of being able to say much about producer behavior and supply. To say more, we will need some additional assumptions. Some of these assumptions will always be necessary, others we will need only when they are convenient.

The first three assumptions are ubiquitous:

A1 The production possibilities set is *Nonempty*: **PPS** $\neq \emptyset$.

A2 The production possibilities set is *Closed*.

A3 The production possibilities set satisfies *Free Disposal*: If **y** $\in$ **PPS** and **y'** $\leq$ **y**, then **y'** $\in$ **PPS**.

The first of these assumptions simply says that there is something that can be produced, so there is work for production economists. The second is technical and in a sense implies **PPS** contains its boundary, so we usually use max and min instead of sup and inf in our analysis. The third has more economic content. It says we can always produce less output with the same inputs or we can use more inputs to produce the same output. It is similar to the more is better assumption in consumer theory. When outputs and inputs are clearly delineated, this assumption can be stated separately for outputs and inputs:

A3.IS *Strong Monotonicity/Free Disposal* of inputs: If $\mathbf{z} \in \mathbf{IRS}(\mathbf{q})$ and $\mathbf{z}' \geq \mathbf{z}$, then $\mathbf{z}' \in \mathbf{IRS}(\mathbf{q})$.

A3.OS *Strong Monotonicity/Free Disposal* of output: If $\mathbf{q} \in \mathbf{FOS}(\mathbf{z})$ and $\mathbf{q} \geq \mathbf{q}'$, $\mathbf{q}' \in \mathbf{FOS}(\mathbf{z})$.

There are also weak versions of these assumptions that are often employed instead:

A3.IW *Weak Monotonicity/Free Disposal* of inputs: If $\mathbf{z} \in \mathbf{IRS}(\mathbf{q})$ and $\theta \geq 1$, then $\theta\mathbf{z} \in \mathbf{IRS}(\mathbf{q})$.

A3.OW *Weak Monotonicity/Free Disposal* of output: If $\mathbf{q} \in \mathbf{FOS}(\mathbf{z})$ and $1 \geq \theta > 0$, then $\theta\mathbf{q} \in \mathbf{FOS}(\mathbf{z})$.

The strong free disposal assumptions simply say that you can always produce less of one or more outputs with the same amount of inputs or the same amount of output with more of one or more inputs. The weak free disposal assumptions say you can always produce proportionally less output with the same amount of inputs or the same amount of output with proportionally more inputs. If $\mathbf{IRS}(\mathbf{q})$ or $\mathbf{FOS}(\mathbf{z})$ satisfies strong free disposal, it will also satisfy weak free disposal. Compared to A3, A3.IS and A3.OS are more restrictive because we must assume which commodities are outputs and which are inputs, but less restrictive because outputs can satisfy strong free disposal while inputs do not or inputs can satisfy strong free disposal while outputs do not.

With weak free disposal, it is possible to formally establish some additional properties of distance functions that were alluded to above:

Inputs satisfy weak free disposal if and only if $\mathbf{IRS}(\mathbf{q}) = \{\mathbf{z}: D_I(\mathbf{q}, \mathbf{z}) \geq 1\}$.

Outputs satisfy weak free disposal if and only if $\mathbf{FOS}(\mathbf{z}) = \{\mathbf{q}: D_O(\mathbf{q},\mathbf{z}) \leq 1\}$.

Therefore, with weak free disposal, we can use the input and output distance functions to characterize feasible production. To give ourselves some confidence that these results are true, let us try to formally show the first one.

First suppose that inputs satisfy weak free disposal, but there exists a $\mathbf{z}$' such that (i) $\mathbf{z}' \in \mathbf{IRS}(\mathbf{q})$ and $D_I(\mathbf{q},\mathbf{z}') < 1$ or (ii) $\mathbf{z}' \notin \mathbf{IRS}(\mathbf{q})$ and $D_I(\mathbf{q},\mathbf{z}') \geq 1$. Consider (i). By definition of the input distance function, $\frac{\mathbf{z}'}{\delta} \in \mathbf{IRS}(\mathbf{q})$ for all $1 > D_I(\mathbf{q},\mathbf{z}') \geq \delta$, which is a contradiction because $\mathbf{z}' \in \mathbf{IRS}(\mathbf{q})$ for $\delta = 1$ by assumption. For (ii), note that by the definition of the input distance function $\frac{\mathbf{z}'}{D_I(\mathbf{q},\mathbf{z}')} \in \mathbf{IRS}(\mathbf{q})$, while weak free disposal then also implies $\frac{\mathbf{z}'}{\delta} \in \mathbf{IRS}(\mathbf{q})$ for all $D_I(\mathbf{q},\mathbf{z}') \geq \delta$. By assumption, $D_I(\mathbf{q},\mathbf{z}') \geq 1$, which implies $\frac{\mathbf{z}'}{\delta} \in \mathbf{IRS}(\mathbf{q})$ for all $D_I(\mathbf{q},\mathbf{z}') \geq \delta \geq 1$. By weak free disposal of inputs, if $\frac{\mathbf{z}'}{\delta} \in \mathbf{IRS}(\mathbf{q})$ and $\theta \geq 1$, then $\theta\frac{\mathbf{z}'}{\delta} \in \mathbf{IRS}(\mathbf{q})$. Therefore, choosing $\theta = \delta \geq 1$ yields the contradiction that $\theta\frac{\mathbf{z}'}{\delta} = \mathbf{z}' \in \mathbf{IRS}(\mathbf{q})$.

Now suppose $\mathbf{IRS}(\mathbf{q}) = \{\mathbf{z}: D_I(\mathbf{q},\mathbf{z}) \geq 1\}$. Given $\mathbf{z} \in \mathbf{IRS}(\mathbf{q})$, $D_I(\mathbf{q},\mathbf{z}) \geq 1$. By the homogeneity property of input distance functions, $D_I(\mathbf{q}, \theta\mathbf{z}) = \theta D_I(\mathbf{q},\mathbf{z})$ for all $\theta > 0$. Then for all $\theta \geq 1$, $D_I(\mathbf{q},\mathbf{z}) \geq 1$ implies $D_I(\mathbf{q}, \theta\mathbf{z}) = \theta D_I(\mathbf{q},\mathbf{z}) \geq 1$. For $D_I(\mathbf{q}, \theta\mathbf{z}) \geq 1$, $\theta\mathbf{z} \in \mathbf{IRS}(\mathbf{q})$. Therefore, if $\mathbf{z} \in \mathbf{IRS}(\mathbf{q})$ and $\theta \geq 1$, then $\theta\mathbf{z} \in \mathbf{IRS}(\mathbf{q})$, which is the definition of weak free disposal for inputs.

We can actually go even further to say:

If inputs satisfy weak free disposal, then $\mathbf{ISQ}(\mathbf{q}) = \{\mathbf{z}: D_I(\mathbf{q},\mathbf{z}) = 1\}$.

If outputs satisfy weak free disposal, then $\mathbf{PPF}(\mathbf{z}) = \{\mathbf{q}: D_O(\mathbf{q},\mathbf{z}) = 1\}$.

Therefore, we can characterize efficient production by setting the input distance function equal to one when inputs satisfy weak free disposal or by setting the output distance function equal to one if outputs satisfy weak free disposal.

The next assumption can simplify our lives.

A4 The production possibilities set is *convex*: For all $\mathbf{y}$, $\mathbf{y}' \in$ **PPS** and all $\alpha \in [0, 1]$, $\alpha\mathbf{y} + (1 - \alpha)\mathbf{y}' \in$ **PPS**.

This convexity assumption is analogous to the convexity assumptions in consumer theory. Intuitively, the assumption implies that balanced input combinations will tend to be more productive than unbalance ones, so it is more than just an assumption of convenience. It drives the notion of diminishing marginal returns. Again, if we know what outputs and inputs are, we can frame this assumption more specifically and to varying degrees:

A4.I The input requirement set is *convex*: For all $\mathbf{z}$, $\mathbf{z}' \in \mathbf{IRS}(\mathbf{q})$ and all $\alpha \in [0, 1]$, $\alpha\mathbf{z} + (1 - \alpha)\mathbf{z}' \in \mathbf{IRS}(\mathbf{q})$.

A4.O The feasible output set is *convex*: For all $\mathbf{q}$, $\mathbf{q}' \in \mathbf{FOS}(\mathbf{z})$ and all $\alpha \in [0, 1]$, $\alpha\mathbf{q} + (1 - \alpha)\mathbf{q}' \in \mathbf{FOS}(\mathbf{z})$.

A4.IS The input requirement set is *strictly convex*: For all $\mathbf{z}$, $\mathbf{z}' \in \mathbf{IRS}(\mathbf{q})$, $\mathbf{z} \neq \mathbf{z}'$ and all $\alpha \in (0, 1)$, $\alpha\mathbf{z} + (1 - \alpha)\mathbf{z}' \in \mathbf{IRS}(\mathbf{q})$ and $\alpha\mathbf{z} + (1 - \alpha)\mathbf{z}' \notin \mathbf{ISQ}(\mathbf{q})$.

A4.OS The feasible output set is *strictly convex*: For all $\mathbf{q}$, $\mathbf{q}' \in \mathbf{FOS}(\mathbf{z})$, $\mathbf{q} \neq \mathbf{q}'$ and all $\alpha \in (0, 1)$, $\alpha\mathbf{q} + (1 - \alpha)\mathbf{q}' \in \mathbf{FOS}(\mathbf{z})$ and $\alpha\mathbf{q} + (1 - \alpha)\mathbf{q}' \notin \mathbf{PPF}(\mathbf{z})$.

Assumptions A4.I and A4.O are the same idea as A4, they are just stated separately for outputs and inputs, which makes them less general because we have to assume which commodities are outputs and which are inputs. However, they are again more general in the sense that the **PPS** need not be convex when A4.I and A4.O is convex. A4.I and A4.O are more restrictive than

A4.I and A4.O, but the tradeoff is that we will be able to guarantee unique solutions for the types of optimization problems of interest to us.

With these convexity assumptions, even more can be said about the output and input distance functions:

If **IRS**($\mathbf{q}$) is convex, $D_I(\mathbf{q}, \mathbf{z})$ is a concave function of $\mathbf{z}$.

If **FOS**($\mathbf{z}$) is convex, $D_O(\mathbf{q}, \mathbf{z})$ is a convex function of $\mathbf{q}$.

For the first of these, note that by definition $D_I(\mathbf{q}, \mathbf{z})$ is a concave function of $\mathbf{z}$ if

$$\alpha D_I(\mathbf{q}, \mathbf{z}^1) + (1-\alpha)D_I(\mathbf{q}, \mathbf{z}^2) \le D_I(\mathbf{q}, \alpha\mathbf{z}^1 + (1-\alpha)\mathbf{z}^2)$$

for all $\mathbf{z}^1, \mathbf{z}^2 \in$ **IRS**($\mathbf{q}$) and $\alpha \in [0,1]$. Suppose this is not the case such that there exists an $\alpha^0 \in [0,1]$ where

$$\alpha^0 D_I(\mathbf{q}, \mathbf{z}^1) + (1-\alpha^0)D_I(\mathbf{q}, \mathbf{z}^2) > D_I(\mathbf{q}, \alpha^0\mathbf{z}^1 + (1-\alpha^0)\mathbf{z}^2).$$

First consider $\alpha^0 = 0$ and $\alpha^0 = 1$, which imply the immediate contradictions $D_I(\mathbf{q}, \mathbf{z}^1) > D_I(\mathbf{q}, \mathbf{z}^1)$ and $D_I(\mathbf{q}, \mathbf{z}^2) > D_I(\mathbf{q}, \mathbf{z}^2)$.

Now for $\alpha^0 \in (0,1)$, the homogeneity of inputs implies

$$D_I(\mathbf{q}, \alpha^0\mathbf{z}^1) + D_I(\mathbf{q}, (1-\alpha^0)\mathbf{z}^2) > D_I(\mathbf{q}, \alpha^0\mathbf{z}^1 + (1-\alpha^0)\mathbf{z}^2).$$

Note that $\alpha^0\mathbf{z}^1 + (1-\alpha^0)\mathbf{z}^2 \in$ **IRS**($\mathbf{q}$) for all $\mathbf{z}^1, \mathbf{z}^2 \in$ **IRS**($\mathbf{q}$) by convexity. We also know $\frac{\alpha^0\mathbf{z}^1}{D_I(\mathbf{q},\alpha^0\mathbf{z}^1)}, \frac{(1-\alpha^0)\mathbf{z}^2}{D_I(\mathbf{q},(1-\alpha^0)\mathbf{z}^2)} \in$ **IRS**($\mathbf{q}$) by the definition of the input distance function. Convexity implies $\beta\frac{\alpha^0\mathbf{z}^1}{D_I(\mathbf{q},\alpha^0\mathbf{z}^1)} + (1-\beta)\frac{(1-\alpha^0)\mathbf{z}^2}{D_I(\mathbf{q},(1-\alpha^0)\mathbf{z}^2)} \in \mathbf{IRS}(\mathbf{q})$ for any $\beta \in [0,1]$. Let $\beta = \frac{D_I(\mathbf{q},\alpha^0\mathbf{z}^1)}{D_I(\mathbf{q},\alpha^0\mathbf{z}^1)+D_I(\mathbf{q},(1-\alpha^0)\mathbf{z}^2)}$ such that

$$\beta\frac{\alpha^0\mathbf{z}^1}{D_I(\mathbf{q},\alpha^0\mathbf{z}^1)} + (1-\beta)\frac{(1-\alpha^0)\mathbf{z}^2}{D_I(\mathbf{q},(1-\alpha^0)\mathbf{z}^2)} = \frac{\alpha^0\mathbf{z}^1+(1-\alpha^0)\mathbf{z}^2}{D_I(\mathbf{q},\alpha^0\mathbf{z}^1)+D_I(\mathbf{q},(1-\alpha^0)\mathbf{z}^2)} \in \mathbf{IRS}(\mathbf{q}).$$

By the definition of the input distance function, we then get the contradiction

$$D_I(\mathbf{q},\alpha^0\mathbf{z}^1) + D_I(\mathbf{q},(1-\alpha^0)\mathbf{z}^2) \leq D_I(\mathbf{q},\alpha^0\mathbf{z}^1 + (1-\alpha^0)\mathbf{z}^2).$$

Similar arguments can be made for the second of these.

Some additional assumptions that are often applied based on their intuitive appeal are

A5 *No Free Lunch*: If $\mathbf{y} \in PPS$ and $\mathbf{y} \geq 0$, then $\mathbf{y} = 0$ or if $\mathbf{z} = 0$, then $\mathbf{q} = 0$.

A6 *Feasibility of Inactivity*: $0 \in \mathbf{PPS}$.

A7 *Irreversibility*: If $\mathbf{y} \in \mathbf{PPS}$ and $\mathbf{y} \neq 0$, $-\mathbf{y} \notin \mathbf{PPS}$.

A5 simply says you do not get something from nothing. A6 says we can always choose to do nothing. A7 says that if we produce a thousand shirts from a ton of cotton, but then decide we would rather sell the cotton, we cannot use the thousand shirts to reproduce a ton of cotton.

Above we mentioned that some assumptions may make sense in the long-run, but not in the short-run. A6 is a good example because by our definition of the short-run this assumption can only be true for $\mathbf{y}^f = 0$. That is, the case where we are restricted from producing or using certain commodities.

Most of what we do in this class will rely heavily on assumptions A1, A2, various versions of A3, and various versions of A4.

RETURNS TO SCALE IN PRODUCTION

A fascination with the observation that some industries are characterized by few very large producers, others are characterized by many small producers, and still others are characterized by producers of varying size has resulted in the study of economies of scale and the implications of scale on the production possibilities set. The net outcome of this work is a taxonomy of production possibilities sets in terms of non-increasing, non-decreasing, and constant to returns to scale:

A Production Possibilities Set exhibits *Non-Increasing Returns to Scale* (*NIRS*) if any feasible production vector can be scaled down: if $\mathbf{y} \in \mathbf{PPS}$, then $\alpha\mathbf{y} \in \mathbf{PPS}$ for all $\alpha \in [0, 1]$.

A Production Possibilities Set exhibits *Non-Decreasing Returns to Scale* (*NDRS*) if any feasible production vector can be scaled up: if $\mathbf{y} \in \mathbf{PPS}$, then $\varepsilon\mathbf{y} \in \mathbf{PPS}$ for all $\varepsilon \geq 1$.

A Production Possibilities Set exhibits *Constant Returns to Scale* (*CRS*) if any feasible production vector can be scaled up and down: if $\mathbf{y} \in \mathbf{PPS}$, then $\tau\mathbf{y} \in \mathbf{PPS}$ for all $\tau \geq 0$.

In terms of a production possibilities set with a single output and many inputs, $q = f(\mathbf{z})$, *NIRS* implies

$$\tau q \geq f(\tau\mathbf{z}) \text{ for all } \tau \geq 1,$$

NDRS implies

$$f(\tau\mathbf{z}) \geq \tau q \text{ for all } \tau \geq 1,$$

and *CRS* implies

$$\tau q = f(\tau \mathbf{z}) \text{ for all } \tau > 0.$$

Intuitively, *NIRS*/*NDRS* implies that a proportional increase in inputs will produce less/more than or equal to the same proportional increase in output. *CRS* implies that a proportional increase in inputs leads to the same proportional increase in outputs.

For **PPS**s that exhibit *CRS*, the input and output distance functions will have some additional useful properties:

The **PPS** exhibits *CRS* if and only if the output distance function is homogenous of degree -1 in inputs and the input distance function is homogeneous of degree -1 in outputs: $D_O(\mathbf{q}, \tau\mathbf{z}) = \tau^{-1} D_O(\mathbf{q}, \mathbf{z})$ and $D_I(\tau\mathbf{q}, \mathbf{z}) = \tau^{-1} D_I(\mathbf{q}, \mathbf{z})$ for all $\tau > 0$.

The **PPS** exhibits *CRS* if and only if the output distance function is the inverse of the input distance function: $D_I(\mathbf{q}, \mathbf{z}) = D_O(\mathbf{q}, \mathbf{z})^{-1}$.

The definitions for *NIRS*, *NDRS*, and *CRS* above are all global in nature when in fact there could be lots of interesting **PPS**s that vary in terms of returns to scale locally. The *Elasticity of Scale* provides a way to measure local returns to scale. For **PPS**s with a single output and many inputs, the elasticity of scale is defined as

P3 $$e(\mathbf{z}) = \frac{dln(f(\varepsilon\mathbf{z}))}{dln(\varepsilon)} = \frac{df(\varepsilon\mathbf{z})}{d\varepsilon}\frac{\varepsilon}{f(\varepsilon\mathbf{z})} \text{ where } \varepsilon = 1.$$

Intuitively, the elasticity of scale tells us the percentage change in output for a one percent increase in all factors. Expanding equation P3 yields

P3' $$e(\mathbf{z}) = \sum_{n=1}^{N} \frac{\partial f(\mathbf{z})}{\partial z_n}\frac{z_n}{f(\mathbf{z})} = \sum_{n=1}^{N} e_n(\mathbf{z})$$

where $e_n(\mathbf{z}) = \frac{\partial f(\mathbf{z})}{\partial z_n}\frac{z_n}{f(\mathbf{z})}$ is known as the *Output Elasticity* of input *n*. Therefore, equation P3' tells us that the elasticity of scale equals the sum of the output elasticities of the inputs.

For production possibilities sets with many outputs and many inputs, the elasticity of scale can be defined by appealing to the input and output distance functions. For the input distance function, the elasticity of scale is defined by

P4 $$e_I(\mathbf{q},\mathbf{z}) = \frac{dln\theta}{dln\lambda} = \frac{d\theta}{d\lambda}\frac{\lambda}{\theta} \text{ where } \theta = \lambda = 1 \text{ and } D_I(\theta\mathbf{q},\lambda\mathbf{z}) = 1.$$

Totally, differentiating $D_I(\theta\mathbf{q},\lambda\mathbf{z}) = 1$ with respect to θ and λ implies

$$\frac{d\theta}{d\lambda} = -\frac{\sum_{n=1}^{N}\frac{\partial D_I(\theta\mathbf{q},\lambda\mathbf{z})}{\partial z_n}z_n}{\sum_{m=1}^{M}\frac{\partial D_I(\theta\mathbf{q},\lambda\mathbf{z})}{\partial q_m}q_m}.$$

Since the input distance function is homogeneous of degree one in inputs, we know $\sum_{n=1}^{N}\frac{\partial D_I(\mathbf{q},\mathbf{z})}{\partial z_n}z_n = D_I(\mathbf{q},\mathbf{z}) = 1$. Therefore,

P4' $$e_I(\mathbf{q},\mathbf{z}) = -\frac{1}{\sum_{m=1}^{M}\frac{\partial D_I(\mathbf{q},\mathbf{z})}{\partial q_m}q_m}.$$

For the output distance function, the elasticity of scale is defined by

P5 $$e_O(\mathbf{q},\mathbf{z}) = \frac{dln\theta}{dln\lambda} = \frac{d\theta}{d\lambda}\frac{\lambda}{\theta} \text{ where } \theta = \lambda = 1 \text{ and } D_O(\theta\mathbf{q},\lambda\mathbf{z}) = 1,$$

which after similar gyrations yields

P5' $$e_O(\mathbf{q},\mathbf{z}) = -\sum_{n=1}^{N}\frac{\partial D_O(\mathbf{q},\mathbf{z})}{\partial z_n}z_n.$$

These elasticities tell us the percentage change in outputs given the percentage change in inputs when outputs and inputs change proportionally.

INPUT & OUTPUT SUBSTITUTION POSSIBILITES

As economists, we are always interested in tradeoffs. When we look at how and what producers choose to produce, these tradeoffs become crucial. One important measure of how inputs can be traded off is the *Marginal Rate of Technical Substitution* (*MRTS*), which is also referred to as the *Technical Rate of Substitution*. The *MRTS* measures the amount by which one factor of production must increase in order to maintain the same level of production when another factor decreases assuming technically efficient production. With a single output and many factors, we can totally differentiate $q = f(\mathbf{z})$ with respect to q, z_l, and z_k to get

$$dq = \frac{\partial f(\mathbf{z})}{\partial z_l} dz_l + \frac{\partial f(\mathbf{z})}{\partial z_k} dz_k.$$

Setting $dq = 0$ and solving then yields

P6 $$\frac{dz_l}{dz_k} = -\frac{\frac{\partial f(\mathbf{z})}{\partial z_k}}{\frac{\partial f(\mathbf{z})}{\partial z_l}},$$

which is the slope of an isoquant. The *MRTS* is the slope of an isoquant, though the convention is to drop the minus sign: $MRTS = \frac{\frac{\partial f(\mathbf{z})}{\partial z_k}}{\frac{\partial f(\mathbf{z})}{\partial z_l}}$. Recall that $D_I(\mathbf{q}, \mathbf{z}) = 1$ if $\mathbf{z} \in \mathbf{ISQ}(\mathbf{q})$, so if $D_I(\mathbf{q}, \mathbf{z})$ is differentiable in $\mathbf{z}$, we can also use it to get the *MRTS*:

P7 $$\frac{\partial D_I(\mathbf{q},\mathbf{z})}{\partial z_l} dz_l + \frac{\partial D_I(\mathbf{q},\mathbf{z})}{\partial z_k} dz_k = 0 \text{ or } MRTS = \frac{\frac{\partial D_I(\mathbf{q},\mathbf{z})}{\partial z_k}}{\frac{\partial D_I(\mathbf{q},\mathbf{z})}{\partial z_l}}.$$

We can look at tradeoffs even more generally with the transformation function:

P8 $\frac{\partial T(\mathbf{y})}{\partial y_l} dy_l + \frac{\partial T(\mathbf{y})}{\partial y_k} dy_k = 0$ or $\frac{dy_l}{dy_k} = -\frac{\frac{\partial T(\mathbf{y})}{\partial y_k}}{\frac{\partial T(\mathbf{y})}{\partial y_l}}$.

When y_l and y_k are inputs, we again have the *MRTS*. When y_l and y_k are output, we have what is referred to as the *Marginal Rate of Transformation* (*MRT*). The *MRT* is the amount by which one output must decrease, while another output increases in order to maintain efficient production with the same level of inputs.

There is a common class of functions frequently used by economist because they have particularly nice properties in terms of the *MRTS* and *MRT*.

Homothetic Function: A function that is an increasing monotonic transformation of a function that is homogeneous of degree one: $f(\mathbf{z}) = g(h(\mathbf{z}))$ where $g(\cdot)$ is an increasing monotonic function and $h(\mathbf{z})$ is homogenous of degree 1 in $\mathbf{z}$.

Suppose we have a single output, many input **PPS** that yields the homothetic production function $q = f(\mathbf{z})$, then

(i) If $f(\mathbf{z}^0) = f(\mathbf{z}^1)$, then $f(\alpha\mathbf{z}^0) = f(\alpha\mathbf{z}^1)$ for all $\alpha > 0$ and

(ii) $\frac{\frac{\partial f(\mathbf{z})}{\partial z_k}}{\frac{\partial f(\mathbf{z})}{\partial z_l}} = \frac{\frac{\partial f(\alpha\mathbf{z})}{\partial z_k}}{\frac{\partial f(\alpha\mathbf{z})}{\partial z_l}}$ for all l and k and all $\alpha > 0$.

Intuitively, the *MRTS* is constant along a ray passing through the origin.

Figure 1