## Contents

1 Choice Theory ........................................ 7
   1.1 Transitivity? ....................................... 7
      1.1.1 Guiding question (202 Final 2005, #1) .............. 7
      1.1.2 Solution to guiding question ........................ 8
   1.2 Path invariance ..................................... 11
      1.2.1 Guiding question - MWG 1.D.4 ....................... 11
      1.2.2 Why this is interesting ............................ 11
      1.2.3 Solution to guiding question ....................... 11
   1.3 Stochastic choice ................................... 12
      1.3.1 Guiding question - MWG 1.D.5 ....................... 12
      1.3.2 Solution to guiding question ....................... 13
   1.4 The weak axiom of revealed preference ................ 15
      1.4.1 Guiding question - MWG 1.C.3 ....................... 15
      1.4.2 Solution to the guiding question ................... 15
   1.5 Series, set closure, and free disposal ................ 16
      1.5.1 Guiding problem - MWG 5.B.5 ....................... 16
      1.5.2 Solution to guiding problem ....................... 16

2 Producer Theory I .................................... 19
   2.1 Implicit Function Theorem ........................... 19
      2.1.1 Guiding Question ................................. 19
      2.1.2 Background ....................................... 19
      2.1.3 Solution to guiding question ..................... 21
   2.2 Envelope Theorem (Differential Form) ............... 22
      2.2.1 Guiding Question (MWG 5.C.13) ..................... 22
      2.2.2 Background ....................................... 23
      2.2.3 Solution to guiding question ..................... 26
   2.3 Topkis’ Theorem .................................... 27
      2.3.1 Guiding Question (Spence!) ....................... 27
      2.3.2 Background ....................................... 27
      2.3.3 Solution to guiding question ..................... 28

3 Producer Theory II ................................... 31
   3.1 Kuhn-Tucker Algorithm ............................. 31
      3.1.1 Guiding Question (Luenberger) ..................... 31
      3.1.2 Background ....................................... 32
## CONTENTS

3.1.3 Solution to guiding question .................................................. 34  
3.2 Envelope Theorem (Integral Form) ............................................. 39  
  3.2.1 Guiding Question (Cost-benefit analysis) .............................. 39  
  3.2.2 Background .......................................................................... 39  
  3.2.3 Solution to guiding problem .................................................. 40  

4 Consumer Theory I ..................................................................... 43  
  4.1 Duality .................................................................................... 43  
    4.1.1 Guiding Question (MWG 3.G.16) .......................................... 43  
    4.1.2 Background ........................................................................ 43  
    4.1.3 Solution to guiding question .............................................. 45  
  4.2 Homotheticity ......................................................................... 49  
    4.2.1 Guiding Question .............................................................. 49  
    4.2.2 Background ........................................................................ 49  
    4.2.3 Solution to guiding problem .............................................. 50  
  4.3 Aggregation and the representative consumer ......................... 50  
    4.3.1 Guiding question (MWG 4.B.2) .......................................... 50  
    4.3.2 Background ........................................................................ 50  
    4.3.3 Solution to guiding question .............................................. 52  

5 Consumer Theory II .................................................................. 53  
  5.1 Welfare measures and path dependence .................................... 53  
    5.1.1 Guiding Question (2005 202N Midterm, Question 5) ........... 53  
    5.1.2 The easy part (a and b) ...................................................... 54  
    5.1.3 Background on welfare measures ....................................... 55  
    5.1.4 The harder part (c) ........................................................... 56  
    5.1.5 Background on path dependence and conservative fields ...... 56  
    5.1.6 The hardest part (Extra Credit) .......................................... 59  
  5.2 Nested Optimization and the Le Châtelier Principle .................. 59  
    5.2.1 Guiding Question (June 2004 Comp, Question 1) .................. 59  
    5.2.2 Background ........................................................................ 60  
    5.2.3 Solution to guiding question .............................................. 60  

6 Choice under Uncertainty ......................................................... 67  
  6.1 Measures of risk aversion ....................................................... 67  
    6.1.1 Coefficient of absolute risk aversion .................................... 67  
    6.1.2 Coefficient of relative risk aversion .................................... 69  
  6.2 Method of proof in insurance/portfolio problems ..................... 69  
    6.2.1 Insurance .......................................................................... 69  
    6.2.2 Portfolio ............................................................................ 73  
  6.3 Portfolio theory applied ............................................................ 76  
    6.3.1 Guiding question (#3 from the 2004 202 Final) .................... 76  
    6.3.2 Solution ............................................................................. 77  
  6.4 Precautionary savings .............................................................. 79  
    6.4.1 Guiding question (MWG 6.C.9) .......................................... 79  
    6.4.2 Solution ............................................................................. 80
7 General Equilibrium I 83
  7.1 Hemicontinuity and fixed-point theorems .................................. 83
    7.1.1 Correspondences .................................................. 83
    7.1.2 Hemicontinuity .................................................... 83
    7.1.3 Fixed point theorems .............................................. 85
  7.2 Existence of Walrasian equilibrium ........................................... 88
    7.2.1 Guiding problem (from a past 202 final) ............................. 88
    7.2.2 Notation .............................................................. 88
    7.2.3 Solution to guiding problem ......................................... 89
  7.3 Corner solutions and zero prices in equilibrium problems ............... 90
    7.3.1 Guiding problem (MWG 15.B.9) .................................. 90
    7.3.2 Solution to guiding problem ......................................... 91
  7.4 Cutesy tricks with horrible algebra .......................................... 94
    7.4.1 Guiding problem (MWG 15.B.6) .................................. 94
    7.4.2 Pep talk ............................................................... 94
    7.4.3 Solution to guiding problem ......................................... 95

8 General Equilibrium II 99
  8.1 Continuous GE ............................................................. 99
    8.1.1 Guiding question (#3, June 2005 Comp) ............................. 99
    8.1.2 Solution to guiding question ........................................ 100
  8.2 Market failure due to externality ......................................... 105
    8.2.1 Guiding question (#3, June 2004 Comp) ............................. 105
    8.2.2 Externality ............................................................ 106
    8.2.3 Solution to guiding problem ......................................... 106
  8.3 Market failure due to overproduction ...................................... 108
    8.3.1 Guiding question (MWG 16.G.5) .................................. 108
    8.3.2 Overproduction ....................................................... 109
    8.3.3 Solution to guiding problem ......................................... 109

9 Final Exam Review 117
  9.1 Optimal Durability ....................................................... 117
    9.1.1 Micro comp, June 2005, #1 ....................................... 117
    9.1.2 Solution ............................................................... 118
  9.2 Labor managed firms ...................................................... 120
    9.2.1 Micro comp, June 2005, #1 ....................................... 120
    9.2.2 Solution ............................................................... 121
  9.3 Continuous GE ............................................................. 124
    9.3.1 2005 Econ 202 Final, #3 .......................................... 124
    9.3.2 Solution ............................................................... 124
  9.4 Monopolists, Social Planners, and Taxes .................................. 128
    9.4.1 Question #2, 202 Final 2005 ...................................... 128
    9.4.2 Solution ............................................................... 128

A Duality theorem 133
  A.1 The duality theorem ..................................................... 133
B Euler’s theorem and its converse
   B.1 Euler’s theorem.................................................. 137
      B.1.1 \( f(p) \) is homogeneous of degree 1 \( \Rightarrow f(p) = p \cdot \nabla f(p) \) ...... 137
      B.1.2 \( f(p) \) is homogeneous of degree 1 \( \Leftarrow f(p) = p \cdot \nabla f(p) \) ...... 137
   B.2 Corollary to Euler’s theorem.................................... 138
      B.2.1 \( f(p) \) is homogeneous of degree 1 \( \Rightarrow \nabla f(p) \) is homogeneous of degree 0 .................................................. 138

C Expansion in a basis of mins........................................... 139

D Assortment in continuous general equilibrium...................... 143
Chapter 1

Choice Theory

1.1 Transitivity?

1.1.1 Guiding question (202 Final 2005, #1)

Consider a preference relation $\succsim$ over a finite set $X$ defined by a utility function $u : X \rightarrow \mathbb{R}$ in the following way:

$$a \succsim b \iff u(a) \geq u(b) - 1 \quad (1.1)$$

(The interpretation is that $a \succsim b$ as long as the utility improvement of $b$ over $a$ is “imperceptible”.) Must this preference relation be

(a) reflexive,

(b) complete,

(c) transitive?

The preference relation $\succsim$ defines the corresponding “strictly preferred” relation $\succ$ and “indifference” relation $\sim$ as follows:

$$a \succ b \iff [a \succsim b] \land \neg [b \succsim a] \quad (1.2)$$

$$a \sim b \iff [a \succsim b] \land [b \succsim a] \quad (1.3)$$

(d) Must the strictly preferred relation $\succ$ be transitive?

(e) Must the indifference relation $\sim$ be transitive?

Suppose now that we are instead given a complete preference relation $\succsim$ over a finite set $X$ whose corresponding “strictly preferred” relation is transitive. Can we always find a utility function $u : X \rightarrow \mathbb{R}$ such that $\succsim$ is represented by (1.1)? Prove or disprove with a counterexample, distinguishing between two cases:

(f) $X$ has three elements,

(g) $X$ has more than three elements.
1.1.2 Solution to guiding question

(a) **This is trivially the case.** \( a \succeq a \leftrightarrow u(a) \geq u(a) - 1 \), by definition. So, as long as \( 0 \geq -1 \), we are just fine.

(b) **Again, trivially the case.** For any \( a, b \in X \) we can use \( u(x) \) to determine whether \( a \succeq b \) or \( b \succeq a \) (or both!).

(c) This is where the problem starts getting tricky. “Differencing” in the definition of a preference is the canonical example of intransitive preference. To see this, consider where the definitions lead us

\[
\begin{align*}
a \succeq b & \Rightarrow u(a) \geq u(b) - 1 \\
b \succeq c & \Rightarrow u(b) \geq u(c) - 1 \quad \Rightarrow u(b) - 1 \geq u(c) - 2
\end{align*}
\]

But, \( a \succeq c \Leftrightarrow u(a) \geq u(c) - 1 \) which is a more stringent condition than that implied by \( [a \succeq b] \land [b \succeq c] \). As an example, consider the following

\[
\begin{align*}
u(a) &= 0.5 & u(a) &\geq u(b) - 1 \\
u(b) &= 1 & u(b) &\geq u(c) - 1 \\
u(c) &= 1.6 & u(a) &\not\geq u(c) - 1
\end{align*}
\]

Clearly, for arbitrary \( u(x) \), the derived preference \( \succeq \) need not be transitive.

(d) First, let’s consider what the definition for \( \succ \) tells us. In terms of the utility function \( u(x) \), the definition can be rewritten

\[
\begin{align*}
a \succ b & \Leftrightarrow [u(a) \geq u(b) - 1] \land \neg [u(b) \geq u(a) - 1] \\
a \succ b & \Leftrightarrow u(a) > u(b) + 1
\end{align*}
\]

Using our new and simplified definition, we can check to see if transitivity holds

\[
\begin{align*}
a \succ b & \Leftrightarrow u(a) > u(b) + 1 \\
b \succ c & \Leftrightarrow u(b) > u(c) + 1 \quad \Leftrightarrow u(b) + 1 > u(c) + 2
\end{align*}
\]

And \( u(c) + 2 \) is clearly greater than \( u(c) + 1 \). **So, we conclude that \( u(a) > u(c) + 1 \) and that \( \succ \) is transitive.**

(e) We use the same approach to simplify the definition of \( \sim \).

\[
\begin{align*}
a \sim b & \Leftrightarrow [u(a) \geq u(b) - 1] \land [u(b) \geq u(a) - 1] \\
a \sim b & \Leftrightarrow [u(a) + 1 \geq u(b)] \land [u(b) \geq u(a) - 1]
\end{align*}
\]

Hence indifference, in this context, is equivalent to being within one “util” of value. Clearly, this will not be transitive. Consider the following counterexample

\[
\begin{align*}
u(a) &= 0.5 & a &\sim b \\
u(b) &= 1 & b &\sim c \\
u(c) &= 1.6 & a &\not\sim c
\end{align*}
\]

So, \( \sim \) need not be transitive.
(f) We have three elements to consider with three possible relations joining each pair \((\succ, \prec, \sim)\). Hence, we can simply prove the existence of a representative utility by exhaustion of all 27 possibilities (not all 27 will be possible, due to transitivity).

(i) \(x \succ y \succ z\). Transitivity requires that \(x \succ z\) as well. Hence we can easily present this case as
\[
\begin{align*}
    u(x) &= 200 \\
    u(y) &= 100 \\
    u(z) &= 0
\end{align*}
\]

(ii) \(x \succ y \prec z\). Transitivity yields no restriction. Hence, if \(z \sim x\) we can use
\[
\begin{align*}
    u(x) &= 1.2 \\
    u(y) &= 0 \\
    u(z) &= 1.1
\end{align*}
\]
and if \(z \succ x\), we can use
\[
\begin{align*}
    u(x) &= 100 \\
    u(y) &= 0 \\
    u(z) &= 200
\end{align*}
\]
and if \(x \succ z\), we can use
\[
\begin{align*}
    u(x) &= 200 \\
    u(y) &= 0 \\
    u(z) &= 100
\end{align*}
\]

(iii) \(x \succ y \sim z\). Transitivity requires \(z \not\succ x\). If \(z \sim x\), then
\[
\begin{align*}
    u(x) &= 1.1 \\
    u(y) &= 0 \\
    u(z) &= 0.9
\end{align*}
\]
and if \(z \prec x\), then
\[
\begin{align*}
    u(x) &= 0 \\
    u(y) &= 0.9 \\
    u(z) &= 2
\end{align*}
\]

(iv) \(x \prec y \succ z\). Transitivity yields no restriction. If \(z \sim x\), we can use
\[
\begin{align*}
    u(x) &= 0.8 \\
    u(y) &= 2 \\
    u(z) &= 0.9
\end{align*}
\]
If \(z \succ x\), we can use
\[
\begin{align*}
    u(x) &= 0 \\
    u(y) &= 200 \\
    u(z) &= 100
\end{align*}
\]
If \(x \succ z\), we can use
\[
\begin{align*}
    u(x) &= 100 \\
    u(y) &= 200 \\
    u(z) &= 0
\end{align*}
\]

(v) \(x \prec y \prec z\). Transitivity requires \(x \prec z\). We represent this with
\[ u(x) = 0 \]
\[ u(y) = 100 \]
\[ u(z) = 200 \]

(vi) \( x \prec y \sim z \). Transitivity requires \( z \not\prec x \). If \( z \sim x \), the we can use
\[ u(x) = 0.9 \]
\[ u(y) = 2 \]
\[ u(z) = 1.1 \]
and if \( x \prec z \) then we can use
\[ u(x) = 0.9 \]
\[ u(y) = 2 \]
\[ u(z) = 2.5 \]

(vii) \( x \sim y \succ z \). Transitivity requires \( z \not\succ x \). If \( z \sim x \) then
\[ u(x) = 1.1 \]
\[ u(y) = 2 \]
\[ u(z) = 0.9 \]
and if \( z \prec x \), then we can use
\[ u(x) = 2.2 \]
\[ u(y) = 2 \]
\[ u(z) = 0.9 \]

(viii) \( x \sim y \prec z \). Transitivity requires \( z \not\prec x \). If \( z \sim x \) then
\[ u(x) = 2.2 \]
\[ u(y) = 2 \]
\[ u(z) = 3.1 \]
and if \( z \succ x \), then we can use
\[ u(x) = 1.1 \]
\[ u(y) = 2 \]
\[ u(z) = 3.1 \]

(ix) \( x \sim y \sim z \). Transitivity requires nothing here. If \( x \sim z \), then we use
\[ u(x) = 1 \]
\[ u(y) = 1 \]
\[ u(z) = 1 \]
If \( x \succ z \), then we use
\[ u(x) = 2 \]
\[ u(y) = 1.5 \]
\[ u(z) = 0.9 \]
and if \( z \succ z \) then we use
\[ u(x) = 0.9 \]
\[ u(y) = 1.5 \]
\[ u(z) = 0.9 \]

There were 27 possible orderings, but 8 were ruled out by transitivity. We constructed a utility function for all 19 cases, thus showing that a utility representation that follows the definition exists for any preference over a three element choice set, where the derived “strictly preferred” relation is transitive.\(^1\)

\(^1\)This isn’t so bad though. Once you get the idea, you can fly through this proof. The moral of
(g) Consider the four element choice set \( \{a, b, c, d\} \). Let \( a \prec b \prec c \), and let \( d \) be indifferent to all three. Transitivity only tells us that \( a \prec c \). So, let \( u(a) = 0 \). By definition, \( u(b) > u(a) + 1 \) and \( u(c) > u(b) + 1 \), thus \( u(c) > u(a) + 2 \). Hence \( u(d) \) cannot be within 1 “util” of both \( u(a) \) and \( u(c) \). Hence \( [a \sim d] \land [c \sim b] \) cannot be represented. So, we have found a counter example.

1.2 Path invariance

1.2.1 Guiding question - MWG 1.D.4

Show that a choice structure \((\mathcal{P}, C(\cdot))\) for which a rationalizing preference relation \( \succsim \) exists satisfies the path-invariance property: for every pair \( B_1, B_2 \in \mathcal{P} \) such that \( B_1 \cup B_2 \in \mathcal{P} \) and \( C(B_1) \cup C(B_2) \in \mathcal{P} \), we have \( C(B_1 \cup B_2) = C(C(B_1) \cup C(B_2)) \), that is, the decision problem can be safely subdivided.\(^2\)

1.2.2 Why this is interesting

Say we choose the best item in a set with \( n \) members, where \( n \) is large. To be sure, we will need \( n(n-1) \) comparisons. But, what if we split up the original set into two sets and run a “tournament”, i.e. pick the best in each set, and then compare them to find the best overall. This requires about \( 2 \cdot \frac{n}{2} \left(\frac{n}{2} - 1\right) + 1 \) comparisons. So for large \( n \), we have reduced our computational load from leading order \( n^2 \) to leading order \( \frac{1}{2} n^2 \). Conceivably, we could run tournaments with more subdivisions and reduce the computational load even more. This is the same strategy used in the merge sort sorting algorithm, which reduces computational load from leading order \( n^2 \) for simple sorts (like the bubble sort) to leading order \( n \log n \). So, this problem shows us that a rational person can reduce the computation they must undertake in finding their most preferred choice from a set.\(^3\)

1.2.3 Solution to guiding question

So, the goal of this problem is to show equality between sets. How do we do this? The most common method is to show that each of the two sets is a subset of the other

\[ A = B \iff (A \subseteq B) \land (B \subseteq A) \quad (1.4) \]

or, more succinctly,

\[ x \in A \iff x \in B \quad (1.5) \]

the story is not to be intimidated by a proof by exhaustion. If each case is easy, as in this problem, then a simple-minded bulldozer will easily beat an elegant thinker, especially in a time-constrained situation. And sometimes the bulldozer really is the only way to go. Math isn’t always elegant, regardless of what you have been told in the past.


\(^3\)For further enrichment, consult http://en.wikipedia.org/wiki/Sorting_algorithm.
A good place to start with this problem is to compare the sets to which we want to apply the choice rule: \(B_1 \cup B_2\) and \(C(B_1) \cup C(B_2)\). Since, by definition, \(C(X) \subseteq X\), we find that

\[
C(B_1) \cup C(B_2) \subseteq B_1 \cup B_2
\]

Now, say that \(x \in C(B_1) \cup C(B_2)\) and \(y \in C(B_1) \cup C(B_2)\). Since \(x\) is a maximizer over a superset of \(C(B_1) \cup C(B_2)\), we see that \(x\) is chosen out of a set that contains \(y\), so \(x \succeq y\). So, if we can show that \(x \in C(B_1) \cup C(B_2)\), then we will also have shown that \(x \in C(C(B_1) \cup C(B_2))\).

To prove this fact, let’s assume the opposite, that \(x \notin C(B_1) \cup C(B_2)\). This means that \(x\) is the \(\succeq\)-maximal element in neither \(B_1\) nor \(B_2\). So, \(\exists a \in B_1, b \in B_2\) s.t. \(a \succ x\) and \(b \succ x\). But, if this were true, then \(x\) could not be \(\succeq\)-maximal in \(B_1 \cup B_2\), since both \(a\) and \(b\) are in that set. Hence we arrive at \(x \notin C(B_1 \cup B_2)\), a contradiction. So, if \(\succeq\) rationalizes the choice rule \(C(\cdot)\), then \(x \in C(B_1 \cup B_2) \implies x \in C(B_1) \cup C(B_2)\).

And, if \(\forall y\) in that set, \(x \succeq y\), then we must conclude that \(x \in C(C(B_1) \cup C(B_2))\), or \(C(B_1 \cup B_2) \subseteq C(C(B_1) \cup C(B_2))\).

So, what about the other direction of inclusion? Consider \(x \in C(C(B_1) \cup C(B_2))\) and \(y \in B_1 \cup B_2\). How can we show that \(x \succeq y\)? If \(y \in B_1\), then \(\exists z \in C(B_1)\) such that \(z \succeq y\) (even if that \(z\) happens to be \(y\)). Now, since \(x \in C(C(B_1) \cup C(B_2))\), we know that \(x \succeq z\). Hence, by transitivity, \(x \succeq y\). The same logic works if \(y \in B_2\).

Clearly, by the definition of \(C(\cdot)\), \(x\) must be an element of \(B_1 \cup B_2\). So, \(x \succeq y\), \(\forall y \in B_1 \cup B_2\). So we have shown that \(x \in C(B_1 \cup B_2)\) or that \(C(C(B_1) \cup C(B_2)) \subseteq C(B_1 \cup B_2)\).

Thus we have shown that \(C(C(B_1) \cup C(B_2)) = C(B_1 \cup B_2)\).

### 1.3 Stochastic choice

#### 1.3.1 Guiding question - MWG 1.D.5

Let \(X = \{x, y, z\}\) and \(\mathcal{B} = \{\{x, y\}, \{y, z\}, \{z, x\}\}\). Suppose that choice is now stochastic in the sense that, for every \(B \in \mathcal{B}\), \(C(B)\) is a frequency distribution over alternatives in \(B\). For example, if \(B = \{x, y\}\), we write \(C(B) = (C_x(B), C_y(B))\), where \(C_x(B)\) and \(C_y(B)\) are nonnegative numbers with \(C_x(B) + C_y(B) = 1\). We say that the stochastic choice function \(C(\cdot)\) can be rationalized by preferences if we can find a probability distribution \(Pr\) over the six possible (strict) preferences relations of \(X\) such that for every \(B \in \mathcal{B}\), \(C(B)\) is precisely the frequency of choices induced by \(Pr\). For example, if \(B = \{x, y\}\), then \(C_x(B) = Pr(\{\succ: x \succ y\})\). This concept originates in Thurstone (1927)\(^4\), and it is of considerable econometric

interest (indeed, it provides a theory for the error term in observable choice).

(a) Show that the stochastic choice function \( C(\{x, y\}) = C(\{y, z\}) = C(\{z, x\}) = (\frac{1}{2}, \frac{1}{2}) \) can be rationalized by preferences.

(b) Show that the stochastic choice function \( C(\{x, y\}) = C(\{y, z\}) = C(\{z, x\}) = (\frac{1}{4}, \frac{3}{4}) \) is not rationalizable by preferences.

(c) Determine the \( 0 < \alpha < 1 \) at which \( C(\{x, y\}) = C(\{y, z\}) = C(\{z, x\}) = (\alpha, 1 - \alpha) \) switches from rationalizable to nonrationalizable.

1.3.2 Solution to guiding question

(a) So, for any two elements of \( X \), we want one to be preferred half the time and the other to be preferred the other half of the time. A simple appeal to symmetry indicates that the only way to do this is to weight all 6 possible preferences with probability \( \frac{1}{6} \). This is easily verified as correct.

(b) So consider the statements \( x \succ y \), \( y \succ z \), and \( z \succ x \). The probability of each of these statements being true is \( \frac{1}{4} \), hence the probability of at least one of these statements being true is at most \( \frac{3}{4} \). So, what is the probability of none of these statements being true? Note that any two of these statements, when transitivity is applied to them, contradict the third (since all preferences are strict, \( \neg [x \succ y] \iff y \succ x \)). For instance, \( [x < y] \land [y < z] \Rightarrow [x < z] \), which contradicts the third statement, \( z \succ x \). Due to this cyclic property, if any two of the statements is contradicted, the third must be true. Hence, the probability of none of the statements being true is zero! So the probability of at least one statement being true is \( 1 > \frac{3}{4} \). Hence, this stochastic choice function if not rationalizable by preferences.

(c) So, extending the logic of the previous part, we find that \( \alpha \geq \frac{1}{3} \) is necessary for rationalizability. Is this all? What if we instead had considered the opposite statements from those above: \( x \prec y \), \( y \prec z \), and \( z \prec x \). Again, at least one will hold, so it will also be necessary that \( 3(1 - \alpha) \geq 1 \Rightarrow \alpha \leq \frac{2}{3} \). Hence, it is clearly necessary that \( \alpha \in [\frac{1}{3}, \frac{2}{3}] \). But, is this a sufficient condition? How can we construct the required \( Pr \)? Assign the following probabilities to the following orderings

\[
\begin{align*}
a & \quad x \succ y \succ z \\
b & \quad y \succ z \succ z \\
c & \quad z \succ x \succ y \\
d & \quad x \succ z \succ y \\
e & \quad z \succ y \succ x \\
f & \quad y \succ x \succ z 
\end{align*}
\]

Then, any set of probabilities that meets the following set of equations will
yield the desired stochastic choice function

\[
\begin{align*}
  a + c + d &= \alpha \\
  a + b + c + d + e + f &= 1 \\
  a + b + f &= \alpha \\
  b + c + e &= \alpha
\end{align*}
\]

Note that we could stop here if we wanted. 4 linear equations in 6 unknowns is an underdetermined system. So long as none of the equations contradict each other, we are assured of a solution. Just for fun though, we will press on, so that we can confirm that we have indeed derived a good answer.\(^5\) Solving for \(\{a, b, c, d\}\), we find

\[
\begin{align*}
  a &= e + 2\alpha - 1 \\
  b &= 1 - \alpha - e - f \\
  c &= 2\alpha - 1 + f \\
  d &= 2 - 3\alpha - e - f
\end{align*}
\]

(As a refresher, you might consider working this out using matrices and the Gauss-Jordan rules. You are not allowed to use a calculator on the comps, and it is not beyond the realm of possibility that you would be asked to invert a 4 by 4 matrix.) Now, what might work out for \(e\) and \(f\)? The uniform distribution will obviously not work. Usually, the big distinction in these sorts of transitivity-centered proofs is whether the ordering is a cyclic or anti-cyclic permutation of \(x, y, z\). Assuming that this distinction will hold, and going for the simplest, most symmetric solution that isn’t just uniform, let’s try to assign the same probability to all cyclic permutations of \(x, y, z\) and another probability to all anticyclic permutations. This gives us \(a = b = c\) and \(d = e = f\). Using the equation for \(d\), we solve \(d = 2 - 3\alpha - 2d\) which yields \(d = \frac{2}{3} - \alpha\). Plugging this in to the equation for \(c\), we find \(c = \alpha - \frac{1}{3}\), which is the same result we get when we plug in the equations for \(a\) and \(b\). So, if we set \(a = b = c = \alpha - \frac{1}{3}\) and \(d = e = f = \frac{2}{3} - \alpha\) in \(Pr\), then we can construct the desired stochastic choice function. \textbf{Hence, \(\alpha \in \left[\frac{1}{3}, \frac{2}{3}\right]\) is a both necessary and sufficient condition for} \(C(\{x, y\}) = C(\{y, z\}) = C(\{z, x\}) = (\alpha, 1 - \alpha)\) to be rationalizable.

\(^5\)I will give you the moral up front. There are three ways to do a math problem: the easy way, the hard way, and the Navy way. The easy way is to look at the solution key, though this is a right reserved for professors and TAs. The hard way is to calculate the answer from first principles in a logical progression. The Navy way is to guess the answer and show that it is correct. If you know the right answer, you don’t have to have a calculation to show where it came from. Just write it down and prove that it’s right. For instance, when asked for a root of the quadratic \(x^2 - 3x + 2 = 0\), the hard way is to derive a root from the quadratic equation. The Navy way is to simply say 1, because \(1^2 - 3 + 2 = 0\). Seriously consider the Navy way. You would be surprised how often the answer is 1, 0, or the uniform distribution.
1.4 The weak axiom of revealed preference

1.4.1 Guiding question - MWG 1.C.3

Suppose that the choice structure \((\mathcal{B}, C(\cdot))\) satisfies the weak axiom. Consider the following two possible revealed preference relations \(\succ^\ast\) and \(\succ^{**}\):

\[
x \succ^\ast y \iff \text{there is some } B \in \mathcal{B} \text{ such that } x, y \in B, x \in C(B), \text{ and } y \notin C(B)
\]

\[
x \succ^{**} y \iff x \succ^\ast y \text{ but not } y \succ^\ast x
\]

where \(\succ^\ast\) is the revealed at-least-as-good relation defined in Definition 1.C.2.

(a) Show that \(\succ^\ast\) and \(\succ^{**}\) give the same relation over \(X\); that is, for any \(x, y \in X\),

\[
x \succ^\ast y \iff x \succ^{**} y.
\]

Is this still true if \((\mathcal{B}, C(\cdot))\) does not satisfy the weak axiom?

(b) Must \(\succ^\ast\) be transitive?

(c) Show that if \(\mathcal{B}\) includes all three-element subsets of \(X\), then \(\succ^\ast\) is transitive.

1.4.2 Solution to the guiding question

(a) \(x \succ^\ast y \Rightarrow x \succ^{**} y\)

Say that \(x \succ^\ast y\). Then, \(\exists B \in \mathcal{B}\) such that \(x, y \in B, x \in C(B),\) and \(y \notin C(B)\).

Thus we can conclude the weaker statement that \(x \succ^\ast y\). Is it the case that \(y \succ^\ast x\)? If it were, then \(\exists B' \in \mathcal{B}\) such that \(x, y \in B', y \in C(B')\). And the WARP states that

\[
\text{If } x, y \in B \cap B', \text{ then } [x \in C(B)] \land [y \in C(B')] \Rightarrow [y \in C(B)] \land [x \in C(B')]
\]

So, \(y \succ^\ast x\) through WARP contradicts our hypothesis that \(x \succ^\ast y\). Thus, we have \([x \succ^\ast y] \land [y/succsim^\ast x] \Rightarrow x \succ^{**} y\).

(b) The relation does not need to be transitive, since we have the freedom to choose \(\mathcal{B}\). For example, what if \(\mathcal{B} = \{\{x, y\}, \{x, y, z\}\}\) and \(C(\{x, y\}) = \{x\} \text{ and } C(\{x, y, z\}) = \{y\}\). Then, we clearly have both \(x \succ^\ast y\) and \(y \succ^\ast x\), though neither statement’s \(\succ^{**}\) counterpart is true.

\[\text{Given a choice structure } (\mathcal{B}, C(\cdot)) \text{ the revealed preference relation } \succ^\ast \text{ is defined by}
\]

\[
x \succ^\ast y \iff \text{there is some } B \in \mathcal{B} \text{ such that } x, y \in B \text{ and } x \in C(B).
\]
and \( C(\{y, z\}) = \{y\} \)? Then, \( x \succ^* y \) and \( y \succ^* z \), but there is no set in \( \mathcal{B} \) that contains both \( x \) and \( z \), so we do not have \( x \succ^* z \). Note that the trick here was limiting the domain of the choice function. Transitivity’s definition requires three elements, hence, if we make it such that we never choose between three elements, transitivity loses its “bite”. Similarly, HARP’s definition requires two elements to be in the intersection of two choice sets. If this never happens, then HARP loses its bite. The moral of the story is that the best way to stop a joke about what happens when Santa and the Easter bunny are sitting in a bar together is to make sure it never happens.

(c) This basically prevents us from using the trick we used to construct the counterexample in (b). Say we have \( x, y, z \in X \). Assume \( x \succ^* y \) and \( y \succ^* z \). By part (a) and the WARP, this is equivalent to \( x \succ^{**} y \) and \( y \succ^{**} z \). And we are assured that \( \{x, y, z\} \in \mathcal{B} \). Hence, by definition of \( \succ^{**} \), we assert that neither \( y \succeq^* x \) nor \( z \succeq^* y \) is true. Since \( \succeq^* \) rationalizes \( (\mathcal{B}, C(\cdot)) \), it must be the case that \( y, z \notin C(\{x, y, z\}) \). It is axiomatic that choice sets are not empty, so all we have left to choose is \( C(\{x, y, z\}) = x \). Hence, \( x \succ^* z \) and we have proven transitivity.

1.5 Series, set closure, and free disposal

1.5.1 Guiding problem - MWG 5.B.5

Show that if \( Y \) is closed and convex, and \( -\mathbb{R}_+^L \subset Y \), then free disposal holds.

1.5.2 Solution to guiding problem

First, consider why we need closed and convex. Can you think of counterexamples where free disposal doesn’t hold if we were to weaken our conditions? This sort of thought experiment often helps us to see how the proof should work.

Say we have \( y \in Y \) and \( v \in -\mathbb{R}_+^L \). We need to show that \( y + v \in Y \). Since the definition of closure involves series, we consider \( n \in \mathbb{R}_+^L \). Since \( -\mathbb{R}_+^L \subset Y \) and \( nv \in -\mathbb{R}_+^L \), we know that \( nv \in Y \). Since \( Y \) is convex, we assert

\[
(1 - \frac{1}{n})y + \frac{1}{n}(nv) = (1 - \frac{1}{n})y + v \in Y
\]

And since \( Y \) is closed, it contains its limit points, so

\[
\lim_{n \to \infty} (1 - \frac{1}{n})y + v \in Y \Rightarrow y + v \in Y
\]

Thus \( Y \) has free disposal.
Figure 1.1: Contradictions for (a) non-convex sets and (b) non-closed sets
Chapter 2

Producer Theory I

2.1 Implicit Function Theorem

2.1.1 Guiding Question

Suppose $f(z)$ is a concave production function with $L - 1$ inputs $(z_1, ..., z_{L-1})$. Suppose also that $\frac{\partial f(z)}{\partial z_l} \geq 0$ for all $l$ and $z \geq 0$ and that the matrix $D^2 f(z)$ is negative definite at all $z$. Use the firm’s first-order conditions and the implicit function theorem to prove the following statements:

a) An increase in the output price always increases the profit-maximizing level of output.

b) An increase in the output price increases the demand for some input.

c) An increase in the price of an input leads to a reduction in the demand for the input.

2.1.2 Background

Consider the problem of solving for the variable $\eta$ in terms of the parameter $q$ in the following:

$$\psi[\eta, q] = 0 \quad (2.1)$$

If we knew $\psi$ explicitly, we would simply invert the equation, if it were possible to do so. Generally though, when is this possible?

Instead of answering this very difficult question, let’s respond to an easier one. Say we have an ordered pair $(\overline{\eta}, \overline{q})$ that solves (2.1). When can we expand a solution of form $\eta(q)$ locally about $(\overline{\eta}, \overline{q})$? Before we answer though, we must know what “locally” means in context.

Specifically, we are looking for a linear approximation of the solution that is exact within an infinitessimal neighborhood of our know solution point, $(\overline{\eta}, \overline{q})$. This
kind of solution is called a **first-order Taylor expansion**. In our case, it looks like

$$
\eta(q) \approx \bar{\eta} + \left[ \frac{d\eta}{dq} \right] (q - \bar{q})
$$

(2.2)

So, to acquire the desired solution vis-à-vis (2.2), we need to know \( \frac{d\eta}{dq} \). How might we get it? Through the chain rule!

$$
\psi[\eta(q), q] = 0
$$

(2.3)

$$
\frac{\partial \psi}{\partial \eta} \frac{d\eta}{dq} + \frac{\partial \psi}{\partial q} = 0
$$

(2.4)

$$
\frac{d\eta}{dq} = \frac{-\frac{\partial \psi}{\partial q}}{\frac{\partial \psi}{\partial \eta}}
$$

(2.5)

Given, (2.5), we should have a local solution, right? Could anything go wrong?

We must have

$$
|\frac{\partial \psi}{\partial \eta}| \neq 0 \text{ at } \bar{q}
$$

(2.6)

for our local solution to be defined. We call (2.6) our **local solubility condition**. Wherever (2.6) holds, we have a nice solution for \( \eta(q) \).

Now that we have an idea of how the Implicit Function Theorem works in one-dimension, let’s generalize to a multivariate setting. We are still trying to solve (2.1), but now, instead of scalars, let \( \psi \) and \( \eta \) be \( N \)-dimensional variable vectors, and let \( q \) be an \( M \)-dimensional parameter vector. Again, we apply the chain rule to (2.1) in its multivariate form, yielding

$$
\frac{\partial \psi(\eta, q)}{\partial \eta} \frac{\partial \eta(q)}{\partial q} + \frac{\partial \psi(\eta, q)}{\partial q} = 0
$$

(2.7)

$$
\frac{\partial \eta(q)}{\partial q} = -\left[ \frac{\partial \psi(\eta, q)}{\partial \eta} \right]^{-1} . \frac{\partial \psi(\eta, q)}{\partial q}
$$

(2.8)

Hence, our generalized local solubility condition is given by

$$
\text{det} \left[ \frac{\partial \psi(\eta, q)}{\partial \eta} \right] \neq 0 \text{ at } q
$$

(2.9)

Wherever (2.9) holds, we have a local solution for \( \eta(q) \). In other words, the implicit function defined by (2.1) exists. What’s more, we have a convenient formula for \( \partial \eta/\partial q \). So, now we can state the Implicit Function Theorem formally.

**Implicit Function Theorem.** A local function \( \eta(q) \) that solves (2.1) exists at all \( q \) where (2.9) holds. This local solution is given by (2.2), and its derivatives are given by (2.8).

Note that this only works if \( \dim \psi = \dim \eta \), since inverses and determinants only exist for square matrices. If \( \dim \psi > \dim \eta \), then the system is overdetermined, and we cannot guarantee a solution. If \( \dim \psi < \dim \eta \), then the system is underdetermined, and we must consider some of our \( \eta \) and free parameters like the \( q \).
2.1.3 Solution to guiding question

Now that we have introduced the relevant theory, we can solve the guiding problem. The firm’s objective is given by

\[ \Pi = p \cdot f(z) - w \cdot z \quad (2.10) \]

The associated first-order condition is thus

\[ p \cdot \frac{\partial f(z)}{\partial z} - w = 0 \quad (2.11) \]

Now, how can we relate this problem to the theory set up in the previous section? In other words, what corresponds to \( \eta, q, \) and \( \psi \)?

\[ \eta \text{ is } z(p,w) \]
\[ q \text{ is } (p,w) \]
\[ \psi \text{ is } p \cdot \frac{\partial f(z)}{\partial z} - w \]

Let’s begin by ensuring that we meet our local solubility condition, (2.9).

\[ \det \left[ \frac{\partial \psi(\eta,q)}{\partial \eta} \right] = \det \left[ p \frac{\partial^2 f(z)}{\partial z^2} \right] \quad (2.12) \]

Now, since \( f(z) \) is concave, its Hessian must be negative semi-definite; in fact, we are told that it is negative definite in the problem. Hence, the right-hand side of (2.12) is non-zero everywhere. Why? \(^1\)

So, we can use the Implicit Function Theorem to calculate derivatives of the factor demands that are implicitly defined by (2.11).

\[ \frac{\partial z(p,w)}{\partial (p,w)} = -\frac{1}{p} \left[ \frac{\partial^2 f(z)}{\partial z^2} \right]^{-1} \frac{\partial}{\partial (p,w)} \left[ p \frac{\partial f(z)}{\partial z} - w \right] \quad (2.13) \]

What do we do with this mess? The notation is really dense. Perhaps expanding it a bit will help:

\[
\begin{pmatrix}
\frac{\partial z_1}{\partial p} & \frac{\partial z_1}{\partial w_1} & \cdots & \frac{\partial z_1}{\partial w_N} \\
\vdots & \vdots & \ddots & \vdots \\
\frac{\partial z_N}{\partial p} & \frac{\partial z_N}{\partial w_1} & \cdots & \frac{\partial z_N}{\partial w_N}
\end{pmatrix} = -\frac{1}{p} \left[ \frac{\partial^2 f(z)}{\partial z^2} \right]^{-1} \begin{pmatrix}
\frac{\partial f}{\partial z_1} & -1 & 0 & 0 & \cdots & 0 \\
0 & -1 & 0 & 0 & \cdots & 0 \\
0 & 0 & -1 & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & 0 & \cdots & -1
\end{pmatrix}
\]

or more succinctly

\[
\begin{pmatrix}
\frac{\partial z}{\partial p} & \frac{\partial z}{\partial w}
\end{pmatrix} = -\frac{1}{p} \left[ \frac{\partial^2 f(z)}{\partial z^2} \right]^{-1} \begin{pmatrix}
\frac{\partial f}{\partial z}
\end{pmatrix} \quad (L-1) \times 1
\]

\[ \begin{pmatrix}
\frac{\partial f}{\partial z}
\end{pmatrix} \quad (L-1) \times 1
\]

\[ \begin{pmatrix}
-1
\end{pmatrix} \quad (L-1) \times 1
\]

---

\(^1\)If a matrix \( A \) were singular, then there would exist a non-zero vector \( x \) such that \( A \cdot x = 0 \). Then we would necessarily have, for that \( x \), \( x' A \cdot x = 0 \). Hence a definite matrix cannot be singular, by contradiction.
Thus,
\[
\frac{\partial z}{\partial p} = -\frac{1}{p} \left[ \frac{\partial^2 f}{\partial z^2} \right]^{-1} \frac{\partial f}{\partial z} \quad (2.16)
\]
\[
\frac{\partial z}{\partial w} = \frac{1}{p} \left[ \frac{\partial^2 f}{\partial z^2} \right]^{-1} \quad (2.17)
\]

Finally, the hard part is done! Let’s reap the fruit of our labor and answer the questions:

a) **An increase in the output price always increases the profit-maximizing level of output.**

The output, \(f(z)\) changes with \(p\) as
\[
\frac{df}{dp} = \frac{\partial f}{\partial z} \frac{\partial z}{\partial p} = -\frac{1}{p} \frac{\partial f}{\partial z} \left[ \frac{\partial^2 f}{\partial z^2} \right]^{-1} \frac{\partial f}{\partial z} \quad (2.18)
\]
The inverse of a symmetric (the Hessian is symmetric), negative definite matrix is also negative definite (why?)\(^2\), and thus \(df/dp > 0\).

b) **An increase in the output price increases the demand for some input.**

We are told that \(\partial f/\partial z > 0\). From the first equality in (2.18) and the answer to part a, we must conclude that \(\exists i\) such that \(\partial z_i/\partial p > 0\) (though this is not necessarily true \(\forall i\)).

c) **An increase in the price of an input leads to a reduction in the demand for the input.**

Here, we are clearly concerned with the derivatives given by (2.17); in fact, we are only concerned with the own cost derivatives. Where are these located? Along the diagonal! And since \(\left[ \frac{\partial^2 f}{\partial z^2} \right]^{-1}\) is negative definite, its diagonal entries must be negative (why)?\(^3\) Thus, \(\forall i, \partial z_i/\partial w_i < 0\).

### 2.2 Envelope Theorem (Differential Form)

#### 2.2.1 Guiding Question (MWG 5.C.13B)

A price-taking firm produces output \(q\) from inputs \(z_1\) and \(z_2\) according to a differentiable concave production function \(f(z_1, z_2)\). The price of its output is \(p > 0\), and

\(^1\)Say \(A\) is negative definite and \(A^{-1}\) is not. Then, \(\exists x\) such that \(x' A^{-1} \cdot x \geq 0\). Let \(\xi = A^{-1} \cdot x\). Then \(\xi' = x'(A^{-1})'\). So, \(\xi A \cdot \xi = x'(A^{-1})' AA^{-1} \cdot x = x'(A^{-1})' x\). Hence \(A^{-1}\) is negative definite by contradiction if it is true that \((A^{-1})' A^{-1} = I\). The inverse is defined by \(AA^{-1} = 1\). Transposing both sides, we find \((A^{-1})' A' = 1\). Since \(A\) is symmetric, we then find that \(A^{-1}\) must be as well, by uniqueness of the inverse.

\(^2\)Let \(e_i\) denote the unit vector in the \(i\)-direction, and let \(a_{ij}\) be the \(ij^{th}\) entry in the negative definite matrix \(A\). Then \(e_i A \cdot e_i = a_{ii} < 0\).
the prices of its inputs are \((w_1, w_2) \gg 0\). However, there are two unusual things about this firm. First, rather than maximizing profit, the firm maximizes revenue (the manager wants her firm to have bigger dollar sales than any other). Second, the firm is cash constrained. In particular, it has only \(C\) dollars on hand before production and, as a result, its total expenditures on inputs cannot exceed \(C\).

Suppose one of your econometrician friends tells you that she has used repeated observations of the firm’s revenues under various output prices, input prices, and levels of the financial constraint and has determined that the firm’s revenue level \(R\) can be expressed as the following function of the variables \((p, w_1, w_2, C)\):

\[
R(p, w_1, w_2, C) = p \left[ \gamma + \log C - \alpha \log w_1 - (1 - \alpha) \log w_2 \right]
\]

(\(\gamma\) and \(\alpha\) are scalars whose values she tells you.) What is the firm’s use of output \(z_1\) when prices are \((p, w_1, w_2)\) and it has \(C\) dollars of cash on hand?

### 2.2.2 Background

When given a problem that provides things like production functions and cost functions, the Implicit Function Theorem helped us. But, what if instead, we are given something like a profit function (i.e. a maximized value instead of an objective function). Then, we must turn to the **Envelope Theorem**. To work our way toward this theorem, let’s consider a simple, unconstrained optimization:

\[
v(q) = \max_x f(x; q)
\]

\(v(q)\) is called the value function of this problem. How does it relate to \(f(x; q)\)? Through the optimal solution (i.e. the argmax), \(x^*(q)\). Hence,

\[
v(q) = f(x^*(q); q)
\]

Now we can use the chain rule to find out about the derivative of the value function:

\[
v'(q) = \left[ \frac{\partial f(x; q)}{\partial x} \right]_{x^*(q)} \cdot \frac{dx^*}{dq} + \left[ \frac{\partial f(x; q)}{\partial q} \right]_{x^*(q)}
\]

And since \(v\) is the solution to an optimization, we can use the first-order necessary conditions \(\left[ \frac{\partial f}{\partial x} \right]_{x^*(q), q} = 0\) to simplify (2.22) to

\[
v'(q) = \left[ \frac{\partial f(x; q)}{\partial q} \right]_{x^*(q)}
\]

In words, (2.23) says that the change in the value function is solely due to the direct change in the objective function, evaluated at the optimum.

The big idea of this derivation is best summarized as

\[
\text{CHAIN RULE} + \text{FONC} = \text{ENVELOPE THEOREM}
\]

---

4If \(x\) is a vector, \(x \gg 0\) implies that all elements of \(x\) are strictly greater than zero.
With this big idea, let’s generalize to a multivariate constrained optimization:

\[ v(q) = \max_x f(x; q) \]
\[ \text{s.t. } g_1(x; q) = b_1 \\
\vdots \\
\text{g}_n(x; q) = b_n \]  

(2.24)

So where do we start? With the chain rule!

\[ \frac{\partial v}{\partial q_i} = \sum_j \frac{\partial f}{\partial x_j} \cdot \frac{\partial x_j}{\partial q_i} + \frac{\partial f}{\partial q_i} \]  

(2.25)

Now what? We need the first order necessary conditions.

\[ \mathcal{L} = f(x; q) + \sum_k \lambda_k (b_k - g_k(x; q)) \]  

(2.26)

\( \left( \frac{\partial \mathcal{L}}{\partial x_i} = 0 \right) \leftrightarrow \frac{\partial f}{\partial x_i} = \sum_k \lambda_k \frac{\partial g_k}{\partial x_i}, \forall i \)  

(2.27)

And finally we use the “big idea”, plugging (2.27) into (2.25) to yield

\[ \frac{\partial v}{\partial q_i} = \sum_j \sum_k \lambda_k \frac{\partial g_k}{\partial x_j} \cdot \frac{\partial x_j}{\partial q_i} + \frac{\partial f}{\partial q_i} \]  

(2.28)

Now what? These \( \frac{\partial g_k}{\partial x_j} \)'s are ugly. Can we simplify them? How? When considering a problem like these, the process of elimination is often helpful. The only thing left to take a derivative of is the set of constraints in (2.24), so that is probably what we need to do.

\[ \sum_j \frac{\partial g_k}{\partial x_j} \frac{\partial x_j}{\partial q_i} = -\frac{\partial g_k}{\partial q_i} \]  

(2.29)

Exchanging the order of summation in (2.28) and substituting from (2.29):

\[ \frac{\partial v}{\partial q_i} = \frac{\partial f}{\partial q_i} - \sum_k \lambda_k \frac{\partial g_k}{\partial q_i} \]  

(2.30)

This is the full envelope theorem. What does it say?

**The change in the value function with respect to a parameter is a combination of the direct change in the objective and the change in the constraint set, as moderated by the Lagrange multipliers**

Now, for a loose end. We used equality constraints in our derivation. What about inequality constraints? In the Kuhn-Tucker setup, constraints that bind have \( \lambda > 0 \) and those that don’t have \( \lambda = 0 \). So only the binding (read: equality) constraints enter our formula in (2.30). Are there any other concerns? What if the set of binding constraints changes with \( q \)? (2.30) is only valid at \( q \) where the set of binding constraints does not change in an open neighborhood.\(^5\)

\(^5\)I assure you that it is almost never an issue in the “real” world
So, how do we remember this complicated formula? The best mnemonic is that the derivative of the Lagrangian related to the optimization problem, evaluated at its arg max is equal to the derivative of the value function for the optimization problem. In short, then envelope theorem tells us that the derivatives of the value function are equal to the derivatives of the Lagrangian. Note that we have to be careful with this mnemonic in the case of a minimization problem. Our standard approach to solving such a problem is to maximize the negative of the objective. This yields the same arg max, but clearly adds a minus sign to the objective. If we go directly from such a Lagrangian, our envelope theorem will be off by factor of $-1$. To illustrate, consider the producer’s cost minimization problem

$$c(q, w) = \min_z w \cdot z \quad \text{s.t.} \quad f(z) \geq q$$  \hspace{1cm} (2.31)

Either of the following Lagrangians is valid.

$$\mathcal{L} = -w \cdot z + \lambda(f(z) - q) \quad \text{or} \quad \mathcal{L} = w \cdot z - \lambda(f(z) - q)$$  \hspace{1cm} (2.32)

Blindly following the rule that “derivatives of the Lagrangian equal derivatives of the value function” then yields two identities:

$$z_l = -\frac{\partial c}{\partial w_l} \quad \text{or} \quad z_l = \frac{\partial c}{\partial w_l}$$  \hspace{1cm} (2.33)

So, which is correct? If you follow the more careful derivation of the envelope theorem that was developed above, you will see that the Lagrangian and envelope identity on the right are correct. Why is this? Intuitively, by negating the objective instead of the Lagrange multipliers, we have negated the value function, leading to a minus sign error. The correct approach is to negate the multipliers. So be careful with your minus signs in applying the envelope theorem to minimization problems.

Finally, why is it called the envelope theorem? To answer, consider the following graphical exercise. Draw the objective function in $q$-space for every possible value of $x$. The “envelope” of these curves is the value function, $v$. So the envelope theorem tells us about the derivative of the envelope function. Note that $v$ doesn’t always run through peak values of the objective in Figure 2.1, but it is always tangent where it touches. So goes the envelope theorem. Now, for completeness, we state the envelope theorem formally.

**Envelope Theorem.** Consider the value function defined by (2.24). Assume it is differentiable. Then, the value function’s derivative is given by (2.30) for all $q$ such that the set of binding constraints does not change in an open neighborhood.

---

6This is non-trivial. Samuelson pointed this out only about fifty years ago when he saw an envelope drawn incorrectly in a 1931 paper by Jacob Viner. The apocryphal anecdote goes that he pointed it out during a lecture and stumped the professor. Another Samuelson story goes that after his dissertation defense, one of his examiners, Joseph Schumpeter, said to the other examiner, Wassily Leontief, “Well, Wassily, have we passed?” The moral of these types of Samuelson stories generally goes like “Samuelson was a smart, smart man.”
2.2.3 Solution to guiding question

So, the Lagrangian in this problem is given by

\[ \mathcal{L} = pf(z_1, z_2) + \lambda(C - w_1z_1 - w_2z_2) \]  \hspace{1cm} (2.34)

How can we use the envelope theorem to pull \( z_1 \) out of the given \( R \)?

\[ \frac{\partial \mathcal{L}}{\partial w_1} = -\lambda z_1 \]  \hspace{1cm} (2.35)

\[ \frac{\partial \mathcal{L}}{\partial C} = \lambda \]  \hspace{1cm} (2.36)

Hence,

\[ z_1 = -\frac{\partial \mathcal{L}/\partial w_1}{\partial \mathcal{L}/\partial C} \]  \hspace{1cm} (2.37)

and by the envelope theorem,

\[ z_1 = -\frac{\partial R/\partial w_1}{\partial R/\partial C} \]  \hspace{1cm} (2.38)

So, we end up with

\[ z_1 = -\frac{\alpha p}{w_1} = \frac{\alpha C}{w_1} \]  \hspace{1cm} (2.39)

In consumer theory, the analogous result to this is called “Roy’s identity”. Another very similar identity is called “Shephard’s Lemma”. Why do we talk about this? Don’t be fooled by the fancy names or bother memorizing these lemmas independently. Learn the envelope theorem and be able to derive them on the fly, and you will have no problems.
2.3 Topkis’ Theorem

2.3.1 Guiding Question (Spence!)

Consider a worker who has to choose an amount of schooling, $x$, and has natural talent of ability $\theta$. Suppose that the worker’s cost of schooling is $c(x, \theta)$.

a) Suppose that $c$ is differentiable and $\partial^2 c(x, \theta)/\partial x \partial \theta < 0$. Interpret this condition economically.

b) Now suppose that firms can observe education, but not ability, and thus offer wages $w(x)$ that depend only on education. The worker’s utility if given by $w(x) - c(x, \theta)$. Is there any wage function $w$ that will induce higher ability workers to choose lower levels of education?

c) Explain why firms might be willing to pay strictly higher wages to workers with higher education, even if education itself is totally unproductive.

2.3.2 Background

Topkis’ theorem is a simple way to make comparative statics conclusions without any unnecessary assumptions. It is guaranteed to be on the comp, so it makes sense to learn it now, so that you will be familiar with it by the time June rolls around.

To start though, we need to define supermodularity. A function is supermodular in $(x, \theta)$ if

$$f(x', \theta') - f(x, \theta) \geq f(x', \theta) - f(x, \theta), \forall x' > x, \theta' > \theta$$  \hspace{1cm} (2.40)

For this reason, supermodularity is sometimes called the increasing differences property. If $f$ is differentiable, then the supermodularity condition reduces to

$$\frac{\partial^2 f}{\partial x \partial \theta} \geq 0$$ \hspace{1cm} (2.41)

Note that these definitions are symmetric in $(x, \theta)$. Supermodularity can also be defined in the language of lattice theory. A poset $(X, \succeq)$ consists of a set $X$ and a partial ordering$^8$ over that set, $\succeq$. Let $\land$ and $\lor$ be operators that we will call the “meet” and the “join”. They are defined$^9$ by

$$x \land y = \inf \{x, y\}$$ \hspace{1cm} (2.42)

$$x \lor y = \sup \{x, y\}$$ \hspace{1cm} (2.43)

$^7$Short for partially ordered set

$^8$An ordering is a partial ordering if it is transitive, reflexive, and antisymmetric ($y \succeq x$ and $x \succeq y$ imply $y = x$). For an example of a poset that does not have a complete ordering, consider the power set of some set, ordered by set inclusion.

$^9$The infimum of a set $A$ is defined in this context by the largest element (relative to $\succeq$) of $X$ that is a lower bound for all elements of $A$. The supremum is defined analogously. So sup and inf are the lowest upper bound and the greatest lower bound, respectively. They are quite different that max and min.
Supermodularity can thus be more generally defined by

\[ f(z \land z') + f(z \lor z') \geq f(z) + f(z'), \forall z, z' \in X \quad (2.44) \]

If \( \succeq \) is the product ordering\(^{10}\), then this definition corresponds to the first definition that we introduced.

Now that this background concerning supermodularity has been introduced, we can properly introduce Topkis’ theorem. Consider the problem

\[ x^*(\theta) \in \arg\max_{x \in D} f(x, \theta) \quad (2.45) \]

where \( f \) is supermodular. Say \((x, \theta)\) and \((x', \theta')\) are two solutions to (2.45). Then, optimization yields the inequalities

\[ f(x, \theta) \geq f(x', \theta) \quad (2.46) \]
\[ f(x', \theta') \geq f(x, \theta') \quad (2.47) \]

Combined, these yield

\[ f(x', \theta') - f(x, \theta') \geq f(x', \theta) - f(x, \theta) \quad (2.48) \]

Hence, if \( \theta' > \theta \), then we must have \( x' \geq x \) to avoid contradicting the supermodularity of \( f \). If \( f \) is supermodular, then \( x^*(\theta) \) must be weakly increasing. There is only one hitch. What if the pairs \((x', \theta)\) and \((x, \theta')\) aren’t elements of the constraint set \( D \)? Then, the inequalities (2.47) don’t necessarily hold. In the terminology of lattice theory, we need the constraint set \( D \) to be closed under \( \land \) and \( \lor \) (“meet” and “join”). Such a set is called a lattice. Hence, the formal statement of Topkis’ Theorem is

**Topkis’ Theorem.** Consider the optimization problem (2.45). If the objective \( f \) is supermodular and the constraint set \( D \) is a lattice, then the \( \arg\max \) \( x^*(\theta) \) is weakly increasing.

Anyways, enough theory. Why should we even bother with Topkis’ theorem? The implicit function theorem seems like an adequate workhorse for comparative statics. First off, Topkis’ theorem is very easy to use. It will also be on the comp, so learning it is definitely to your advantage. The final reason is that Topkis’ theorem makes no assumptions about differentiability. The implicit function theorem requires \( \theta \) to be a continuous parameter. But, what if \( \theta = 1 \) represented monopoly and \( \theta = 0 \) represented perfect competition? We often use Topkis’ theorem to draw conclusions about the comparative statics of discrete choice.

### 2.3.3 Solution to guiding question

a) The marginal cost of education decreases with talent.

---

\(^{10}\) If \( x \succeq y \Rightarrow x_i \geq y_i, \forall i \). Note that this ordering is not complete in the sense used by choice theory. We are concerned with **posets** (partially ordered), not **cosets** (completely ordered).
b) Now the worker faces the problem

$$\max_x w(x) - c(x, \theta)$$  \hspace{1cm} (2.49)

So is the objective supermodular in $(x, \theta)$? Can we just take derivative with wreckless abandon? What if $w(x)$ is discontinuous or has kinks? We will have to do a little fancy footwork to avoid $w(x)$. Plugging our specific objective into the supermodularity definition (2.40):

$$w(x') - c(x', \theta') - w(x) + c(x, \theta) \geq w(x) - c(x', \theta) - w(x) + c(x, \theta)$$  \hspace{1cm} (2.50)

$$c(x, \theta') - c(x', \theta') \geq c(x, \theta) - c(x', \theta) \quad \text{by } \frac{\partial^2 c}{\partial x \partial \theta} < 0$$  \hspace{1cm} (2.51)

So, by Topkis, $x^*$ is non-decreasing in $\theta$ regardless of $w$.

c) This is the standard signaling problem commonly found in markets with asymmetric information. Employers pay for a useless attribute because it acts as a signal of productivity.
Chapter 3  

Producer Theory II  

3.1 Kuhn-Tucker Algorithm  

3.1.1 Guiding Question (Luenberger)  

This problem deals with a common problem of cost accounting. Often, a firm will have several divisions that each produce a separate product, but uses a common input factor. For example, Shell turns crude oil into different grades of oil and gasoline. For accounting purposes, the total cost of the common factor typically is allocated among the divisions. The method of cost allocation is crucial, since input demands are typically determined by the divisions, and the firm wants the aggregate of individual divisions to be optimal for the firm.

Specifically, suppose the firm has $L$ divisions. Each division $i$ produces an output $q_i$, using an independent factor $z_i$ and a common factor $x$. Assume division $i$’s production function $q_i = f_i(z_i, x)$ can be inverted to give $x = g_i(q_i, z_i)$, where $g_i(q_i, z_i)$ is the amount of common factor required to produce $q_i$ units of output $i$ given $z_i$ units of input. Let $w$ be the vector of input prices, and $v$ the price of $x$. The firm’s cost function is

$$c(q, w, v) = \min_{z \geq 0, x \geq 0} \ w \cdot z + vx$$

s.t. $x \geq g_i(q_i, z_i), \forall i$ (3.1)

Let $\lambda_1, \ldots, \lambda_L$ be the Lagrange multipliers associated with these constraints. We propose assigning a price $\lambda_i$ for the use of the common factor by the $i^{th}$ product.

a) Show this method completely allocates the cost of the common factor; i.e. that $v = \sum_i \lambda_i$.

b) Based on this method, construct a decentralized cost minimization problem associated with each output and show that the sum of the divisional cost functions will equal the cost function for the entire firm.

c) Suppose each production function is Leontieff, so $f_i(z_i, z) = \min\{z_i, x\}$. Find the total cost function and determine how the cost of the common factor should be allocated.
3.1.2 Background

In graduate economics, we hear a lot about the Kuhn-Tucker theorem, and we basically take it to be synonymous with constrained maximization. But what is it? How does it help us? Consider the following problem:

$$\min_x f(x; q)$$

s.t. $$g_1(x; q) \leq b_1$$
$$g_2(x; q) \leq b_2$$

(3.2)

Kuhn-Tucker essentially gives us an algorithm for solving this problem. Aside from a few technical issues (like independence of constraints), we start by forming the Lagrangian:

$$L = -f(x; q) + \lambda [b_1 - g_1(x; q)] + \mu [b_2 - g_2(x; q)]$$

(3.3)

From here we take first-order necessary conditions:

$$[x] : \frac{\partial f}{\partial x} = \lambda \frac{\partial g_1}{\partial x} + \mu \frac{\partial g_2}{\partial x}$$

(3.4)

Did we miss any first-order conditions? We are used to considering $$\frac{\partial L}{\partial \lambda} = 0$$ and $$\frac{\partial L}{\partial \mu} = 0$$ as first-order conditions, but this is not strictly correct - they are only necessary if the constraints bind. More generally, the Kuhn-Tucker theorem prescribes complementary slackness conditions:

$$\lambda \cdot (b_1 - g_1(x; q)) = 0$$

(3.5)

$$\mu \cdot (b_2 - g_2(x; q)) = 0$$

(3.6)

In addition to these complementary slackness conditions, Kuhn-Tucker also prescribes multiplier positivity constraints:

$$\lambda, \mu \geq 0$$

(3.7)

Note that if there were an equality constraint in the optimization, then the multiplier would be zero valued (this is easiest to see if we formulate an equality constraint as two weak inequality constraints). Anyways, the long and the short of it is that complementary slackness requires that a multiplier whose constraint does not hold is zero valued.

The intuition of complementary-slackness and multiplier positivity is simple. The multiplier terms in the Lagrangian act as penalties for violating the constraints. The idea of Lagrangian optimization is that, if the penalty per unit of violation is sufficiently high, then the contraint will be obeyed. Positivity constraints ensure that the multiplier terms in the Lagrangian are penalties and not subsidies. If a constraint binds with equality, then that means that a penalty may have had to

\[\text{Note that we have the negative of our objective in the Lagrangian. We choose to formulate minimization as a maximization of the negative of the objective function to keep the signs of the Lagrange multipliers positive. More on this later}\]
be added to keep from breaking the constraint. Hence, we must be able to have a non-zero multiplier. If the constraint does not bind, then we could just as easily solve the optimization without the constraint. This is equivalent to setting the corresponding multiplier to zero in the Lagrangian setup. So, the complementary slackness constraint simply codified this intuition into math.

So, why are these complementary slackness constraints so unfamiliar? We have heard about them but have never really used them. Why are we used to just setting \( \frac{\partial L}{\partial x} = 0 \)?

Because, by and large, we choose constraints that bind. That is why we chose them in the first place. If they don’t bind, their multipliers go to zero, and it is as if they never existed. Also, most problems avoid the slackness conditions because, frankly, they are a pain. So, for the most part, you can just ignore them and plug away with the first-order conditions and the assumption that the constraints bind. But sometimes there are many constraints, and it is not clear which will bind, or if they will bind at all. It is important to get a feel for which constraints will bind. But what if there is no real intuition? That is where the Kuhn-Tucker algorithm sets in.

As a demonstration, let’s solve a familiar problem - the maximization of a function of one variable. In high school calculus, we all learned that the process for this is to

a) Set \( \frac{dy}{dx} \) equal to zero.

b) Check the end points

The Kuhn-Tucker algorithm is an algorithm that generalizes this process to multivariate constrained optimization. It helps us to check the end-points, or corners, for solutions that simple first-order conditions might not find. Let’s illustrate with a more concrete example:

\[
\text{max}_{x \geq 0} -x^3 + 6x^2 - 10x + 5
\]

(3.8)

\[
\mathcal{L} = -x^3 + 6x^2 - 10x + 5 + \mu x
\]

(3.9)

Kuhn-Tucker Step 1 is to check the first-order conditions

\[
\left( \frac{\partial \mathcal{L}}{\partial x} = 0 \right) \iff -3x^2 + 12x - 10 = -\mu
\]

(3.10)

Step 2 is to lay out the slackness and positivity constraints

\[
\mu x = 0; \; \mu \geq 0
\]

(3.11)

Now we can proceed

a) So, say our constraint does not bind. Then \( \mu = 0 \). Solving (3.10) with \( \mu = 0 \), we find \( x = 2 \pm \frac{1}{3} \sqrt{6} \). So these are possible maximizers.

b) Say the constraint does bind. Then \( x = 0 \) and (3.10) gives us \( -10 = -\mu \implies \mu = 10 \). This obeys (3.11), so \( x = 0 \) is a possible maximizer.
As it turns out, the graph looks like

So, it’s a good thing that we checked for sorner solutions!

As an aside, we should note that the first-order conditions are necessary but not sufficient. That is to say they must hold at an optimum, but they can also hold at non-optimal points. When are the first-order conditions sufficient? When the objective is strictly quasiconcave\(^2\) and the constraint set is convex. If these two conditions hold, then the first-order conditions are sufficient to conclude that there is an optimum. Since the objective in the previous problem was not strictly quasi-concave, the first-order conditions led to “spurious” optima.

This was just a simple example to illustrate how the Kuhn-Tucker algorithm simply forces you to “check the end points”, just like your high-school calculus teacher used to do.

### 3.1.3 Solution to guiding question

a) The centralized manager must solve

\[
\begin{align*}
\min_{z_i, x} & \quad \sum_i w_i z_i + v x \\
\text{s.t.} & \quad x \geq g_i(q_i, z_i), \forall i \\
& \quad z_i \geq 0, \forall i \\
& \quad x \geq 0
\end{align*}
\]  

(3.12)

Note that this last constraint is redundant because \(g_i\), by definition, is never less than zero.

\(^2\)Its upper contour sets are concave
So, our Lagrangian is

\[ \mathcal{L} = - \sum_i w_i z_i - vx + \sum_i \lambda_i (x - g_i(q_i, z_i)) + \sum_i \delta_i z_i \]  

(3.13)

Then our first-order conditions are:

\[ [z_i]: \quad - w_i - \lambda_i \frac{\partial g_i}{\partial z_i} + \delta_i = 0, \forall i \]  

(3.14)

\[ [x]: \quad - v + \sum_i \lambda_i = 0 \]  

(3.15)

Our slackness conditions are:

\[ \lambda_i (x - g_i(q_i, z_i)) = 0, \forall i \]  

(3.16)

\[ \delta_i z_i = 0, \forall i \]  

(3.17)

(3.18)

And our positivity conditions are:

\[ \lambda_i, \delta_i \geq 0, \forall i \]  

(3.19)

**Straightaway, we can conclude that** \( v = \sum_i \lambda_i \) **from** (3.15).**

b) How do we proceed from here? To start, what is the decentralized problem? For each division,

\[
\begin{align*}
\min_{z_i, x_i} & \quad w_i z_i + \lambda_i x_i \\
\text{s.t.} & \quad x_i \geq g_i(q_i, z_i) \\
& \quad z_i \geq 0
\end{align*}
\]  

(3.20)

Now what? We solve for the optimality conditions of the decentralized problem, and then show that a solution to the centralized problem must also meet these conditions. In short, we can **plug the centralized optimum into the Kuhn-Tucker conditions for the decentralized problem and show that they hold. This is the standard approach for showing that a market optimum is the same as a planner’s optimum, i.e. that it is Pareto optimal.** So, what are these decentralized optimality conditions? The decentralized Lagrangian is

\[
\mathcal{L}_i = -w_i z_i - \lambda_i x_i + \Lambda_i (x_i - g_i(q_i, z_i)) + \Delta_i z_i
\]  

(3.21)

First-order conditions:

\[ [x_i]: \quad - \lambda_i + \Lambda_i = 0 \]  

(3.22)

\[ [z_i]: \quad - w_i - \Lambda_i \frac{\partial g_i}{\partial z_i} + \Delta_i = 0 \]  

(3.23)

Slackness:

\[ \Delta_i z_i = 0 \]  

(3.24)

\[ \Lambda_i (x_i - g_i(q_i, z_i)) = 0 \]  

(3.25)
Positivity:
\[ \Lambda_i, \Delta_i \geq 0 \] (3.26)

Now what? We can start with a solution to the centralized problem:
\[ x^*, \{z_i^*\}, \{\lambda_i\}, \{\delta_i\} \] (3.27)

We are guaranteed that this solution meets the Kuhn-Tucker conditions. Now what? Let’s try an intuitive guess at the solution for the decentralized problem and see if meets the Kuhn-Tucker conditions for that problem. Set \( x_i = x^* \), \( z_i = z_i^* \), \( \Delta_i = \delta_i \) and \( \Lambda_i = \lambda_i \). Will this fly? The first-order condition for \( x \) in (3.22) is automatically satisfied. The first two positivity constraints in (3.26) are also satisfied. The last of these is satisfied because \( \lambda_i^* \) satisfies (3.19), the corresponding slackness constraint in the simplified problem. (3.22) is trivially satisfied, while (3.23) is guaranteed by the corresponding optimality condition in the centralized problem, (3.14). And positivity, (3.26) also holds.

So, our centralized optimum allocations match those of the decentralized problem.

The firm’s cost function is given by
\[ c(q, w, v) = w \cdot z^* + vx^* = \sum_i w_i z_i^* + vx^* \] (3.28)

and the sum of the decentralized division costs is
\[ c = \sum_i w_i z_i^* + \lambda_i x^* = \sum_i w_i z_i^* + x^* \sum_i \lambda_i = \sum_i w_i z_i^* + vx \] (3.29)

So we have essentially shown the equivalence of the two approaches and their respective cost functions.

c) Now, say all production functions are Leontief: \( f_i(z, x) = \min\{z, x\} \). First, let’s try to work this problem out sans math. What does the form of \( f_i \) tell us about a division’s choice?
\[ z_i = x_i \] (3.30)

And if we constrain a division to produce \( q_i \), then we know
\[ z_i = x_i = q_i \] (3.31)

regardless of internal pricing. So, how much \( x \) does the central manager have to buy for the company?
\[ x = \max_i q_i \] (3.32)

So the total cost is
\[ c = \sum_i w_i q_i + v \max_i q_i \] (3.33)

\(^3\)Note that in this problem, Kuhn-Tucker conditions are sufficient as well as necessary. The objective is linear, which is clear quasiconcave. By assumption, the constraints can be put in the form \( x \geq q_i(q, w_i) \) which is a convex set. Note that this is not always the case if the constraint is expressed in the form \( f(x, z_i) \geq q_i \).
Now, as an exercise, let’s work this same problem out with the Kuhn-Tucker algorithm. To start, what is $g_i(q_i, z_i)$? How do we invert the Leontieff production function?

In this problem, we stated the constraints in terms of the $g_i$’s to make it clear that the constraint set was convex. Here, there is not such a need. The constraint is much more clearly written as $f(x_i, z_i) = \min(x_i, z_i) \geq q_i$. Hence, we can replace one constraint with two:

\begin{align*}
  x &\geq q_i \\
  z_i &\geq q_i
\end{align*}

One of these will be redundant.

Earlier in the problem, we showed that cost is properly and completely allocated if we set the price of the shared factor for each division to the shadow price, $\lambda_i$ of that division’s constraint in the centralized problem. So, let’s look at the centralized problem. Assuming that all the $q_i$ are positive, the positivity constraints on $x$ and $z_i$ become redundant. Thus,

\[
\begin{align*}
  \min_{z,x} & \quad \sum_i w_i z_i + vx \\
  \text{s.t.} & \quad x \geq q_i, \forall i \\
                  & \quad z_i \geq q_i, \forall i
\end{align*}
\]

and its associated Lagrangian is

\[
\mathcal{L} = -\sum_i w_i z_i - vx + \sum_i \lambda_i (x - q_i) + \sum_i \gamma_i (z_i - q_i)
\]

which yields first-order conditions

\[
[x] : \sum_i \lambda_i = v
\]

\[
[z_i] : \gamma_i = w_i
\]

slackness conditions

\[
\lambda_i (x - q_i) = 0, \forall i
\]

\[
\gamma_i (z_i - q_i) = 0, \forall i
\]

and positivity constraints

\[
\lambda_i \geq 0, \forall i
\]

\[
\gamma_i \geq 0, \forall i
\]

Now what do we do?

- $w_i > 0 \Rightarrow \gamma_i > 0 \Rightarrow z_i = q_i$, by (3.41).
- $v > 0 \Rightarrow \exists i \text{ s.t. } \lambda_i > 0 \Rightarrow \exists i \text{ s.t. } x = q_i$ by (3.40).
• But, since $x \geq q_i, \forall i$, $x$ can only equal the highest $q_i$.
• So, $x = \max_i q_i$.

Hence, division $i$’s constraint binds only if $q_i = \max_j q_j$. If this is not the case, then $\lambda_i = 0$.

**Hence the total cost is**

$$c = \sum_i w_i q_i + v \max_i q_i$$  \hspace{1cm} (3.44)

So, our shadow prices take the form

$$\lambda_i = \begin{cases} 
\text{something positive} & q_i = \max_j q_j \\
0 & \text{o.w.}
\end{cases}$$  \hspace{1cm} (3.45)

Along with the extra constraint that $\sum \lambda_i = v$. Now, let’s consider the problem of the individual divisions. They solve

$$\min_{z_i, x_i} w_i z_i + \lambda_i x_i$$

s.t. $x_i \geq q_i, \forall i$

$$z_i \geq q_i, \forall i$$  \hspace{1cm} (3.46)

and its associated Lagrangian is

$$\mathcal{L} = -w_i z_i - \lambda_i x_i + \Lambda_i (x - q_i) + \Gamma_i (z_i - q_i)$$  \hspace{1cm} (3.47)

which yields first-order conditions

$$[x]: \lambda_i = \Lambda_i$$  \hspace{1cm} (3.48)

$$[z_i]: \Gamma_i = w_i$$  \hspace{1cm} (3.49)

slackness conditions

$$\Lambda_i (x_i - q_i) = 0, \forall i$$  \hspace{1cm} (3.50)

$$\Gamma_i (z_i - q_i) = 0, \forall i$$  \hspace{1cm} (3.51)

and positivity constraints

$$\Lambda_i \geq 0, \forall i$$  \hspace{1cm} (3.52)

$$\Gamma_i \geq 0, \forall i$$  \hspace{1cm} (3.53)

So for those divisions where the shadow price $\lambda_i$ is strictly positive, we must have $x_i = q_i = \max_j q_j$. When $\lambda_i$ is zero valued, any value of $x_i$ works, so long as it is greater than $q_i$. So we set $x_i = \max_j q_j$ in this case as well. So the optimal allocation of cost is to charge divisions with $q_i = \max_j q_j$ some positive price for the shared factor and not to charge all other
divisions. What’s more, those positive prices must add up to \( v \) so that the firm completely allocates the cost of the shared factor.

Note that we did this problem generally in the last part, but relied upon the inverse production functions \( g \) to do so. If we return to the generality of part b) and interpret its results in terms of what the \( g_i \)'s are meant to represent, then we come to the same conclusion - that only those divisions whose \( g_i \) constraint binds in the centralized problem will be charged in the decentralized problem. Clearly, upon considering the qualitative interpretation of the \( g_i \)'s as “how much \( x \) is required to produce \( q_i \) given \( z_i \)”, we must see that with a min production function, \( x \) will be determined by \( \max_j q_j \) and that only divisions whose quota attains this max will have their constraints bind.

### 3.2 Envelope Theorem (Integral Form)

#### 3.2.1 Guiding Question (Cost-benefit analysis)

Consider a standard cost-benefit analysis. An agent must decide how much of a good to acquire, \( x \). The benefit of \( x \) is \( u(x) \) and the cost of \( x \) is \( c(x, p) \), where \( p \) is a price. Now, say we know \( c(x, p) \) and that it is differentiable in \( p \). Also, say we observe the agents decision, \( x^*(p) \) for all values of \( p \in (0, p_0] \) where \( p_0 \) is the price at which the agent stops buying any of the good. What can we say about \( u(x) \)? Assume that \( c(0, p) = 0, \forall p \) and \( u(0) = 0 \).

#### 3.2.2 Background

Last week, we introduced the differential form of the envelope theorem. We can also integrate this formula to get an integral form of the envelope theorem. More specifically, consider the problem

\[
\pi(\theta) = \max_x f(x; \theta)
\]  

(3.54)

where \( f \) is absolutely continuous and \( f_\theta \) exists. The differential form of the envelope theorem states

\[
\pi'(\theta) = \frac{\partial f}{\partial \theta}(x(\theta); \theta)
\]  

(3.55)

Integrating we can come to

\[
\int_{\theta_1}^{\theta_2} \pi'(\theta)d\theta = \int_{\theta_1}^{\theta_2} \frac{\partial f}{\partial \theta}(x(\theta); \theta)d\theta
\]  

(3.56)

\[
\pi(\theta_2) = \pi(\theta_1) + \int_{\theta_1}^{\theta_2} \frac{\partial f}{\partial \theta}(x(\theta); \theta)d\theta
\]  

(3.57)

What does this say? At the maximizer, the integral of the full variation with respect to \( \theta \) is equivalent to the integral of the partial variation with respect to \( \theta \).
This is quite similar to the differential envelope theorem, which states that at the maximizer, the full variation of the value function equals the partial variation of the objective function.

Finally, note that this integral form has other names, such as Holmstrom’s Lemma. Again, don’t be fooled by fancy names - just know the envelope theorem in both its forms.

The integral envelope theorem is also important because it is more general that the differential form. It works when the value function is not differentiable. It works when the constraint set has a weird structure. It almost always works. Just for the sake of completeness, I include a very formal statement of integral envelope theorem that was proven in Milgrom and Segal (2002).

**Integral Envelope Theorem (Milgrom-Segal 2002)**

Let $V(t) = \sup_{x \in X} u(x, t)$ be the value function. Let $X^*(t) = \{ x \in X | u(x, t) = V(t) \}$ be the arg max correspondence. Finally let a function $x^*(t)$ be called a selection from the arg max if $x^*(t) \in X^*(t)$ for almost all $t \in [0, 1]$ such that $X^*(0) \neq \emptyset$. Say that $u(x, \cdot) : [0, 1] \mapsto \mathbb{R}$ has the properties

a) There exists a real-valued function $u_2(x, t)$ such that $\forall x \in X$ and every $[a, b] \subset [0, 1], u(x, b) - u(x, a) = \int_a^b u_2(x, s)ds$.

b) There exists an integrable function $b : [0, 1] \mapsto \mathbb{R}_+$ (that is, $\int_0^1 b(s)ds < \infty$) such that $|u_2(x, t)| \leq b(t) \forall x \in X$ and almost all $t \in [0, 1]$.

Further suppose that $X^*(t) \equiv \arg \max_{x \in X} u(x, t) \neq \emptyset$ for almost all $t \in [0, 1]$. Then for any selection $x^*(t)$ from $X^*(t)$,

$$V(t) = u(x^*(t), t) = u(x^*(0), 0) + \int_0^t u_2(x^*(s), s)ds$$  \hspace{1cm} (3.58)

### 3.2.3 Solution to guiding problem

Our agent solves the problem

$$v(p) = \max_x u(x) - c(x, p)$$  \hspace{1cm} (3.59)

Integrating from $p_0$ down, we find that the integral envelope theorem tells us

$$v(p_0) - v(p) = -\int_p^{p_0} c_2(x^*(\pi), \pi)d\pi$$  \hspace{1cm} (3.60)

Since $x^*(p)$ is an arg max, we can also write that

$$v(p) = u(x^*(p)) - c(x^*(p), p)$$  \hspace{1cm} (3.61)
Plugging in to the integral envelope theorem then yields

\[ u(x^*(p_0)) - c(x^*(p_0), p_0) - u(x^*(p)) + c(x^*(p), p) = - \int_{p_0}^{p_0} c_2(x^*(\pi), \pi) d\pi \]  \hspace{1cm} (3.62)

And \( x^*(p_0) = 0, c(0, p) = 0, \) and \( u(0) = 0, \) thus

\[ u(x^*(p)) = c(x^*(p), p) + \int_{p_0}^{p_0} c_2(x^*(\pi), \pi) d\pi \equiv g(p) \]  \hspace{1cm} (3.63)

This must hold for any \( u(x) \) that rationalizes \( x^*(p) \). Does it uniquely define \( u(x) \)? Generally, it does not. For instance, what if \( x^* \) is not monotonic? Say \( x^*(q) = x^*(r), q \neq r \). Then, for a \( u(x) \) to exist that meets our condition, we must require that \( g(q) = g(r) \). If this is not true (and generally, it is not), then there is no \( u(x) \) that rationalizes the given \( x^*(p) \).

What if \( x^*(p) \) has flat segments? \( g(p) \) has to be constant wherever \( x^*(p) \) is, if our condition is to yield a valid \( u(x) \). Let \( x^*(p) = x, \forall p \in [p_1, p_2] \). On this interval, we can divide the integral into two, giving

\[ g(p) = u(x) = c(x, p) + \int_{p_0}^{p_2} c_2(x^*(\pi), \pi) d\pi + \int_{p_2}^{p_0} c_2(x, \pi) d\pi \]  \hspace{1cm} (3.64)

But the last term is now just the integral of a simple derivative. Using the fundamental theorem of calculus, we find

\[ g(p) = c(x, p) + \int_{p_0}^{p_2} c_2(x^*(\pi), \pi) d\pi + c(x, p_2) - c(x, p) \]  \hspace{1cm} (3.65)

which is clearly a constant, since the last two terms cancel. Thus, if \( x^*(p) \) is constant on an interval, so is \( g(p) \). Flat sections of \( x^* \) will give us no problems in finding a \( u \) that rationalizes our integral envelope theorem condition.

Finally, what do we do if \( x^*(p) \) has jumps? This is the same as saying that there are certain \( x \) values that will never be chosen for any \( p \). So, we are free to choose \( u \) so long as it forces the chosen \( x \) at optimum to never be in the jump gaps that \( x^* \) mandates. One easy way to do this is to set \( u(x) = -\infty \) wherever \( x^{*^{-1}} = \emptyset \). Clearly, this is not the only way to rationalize \( x^* \). Hence, if \( x^* \) has jumps, we can find a rationalizing \( u \), but it is not necessarily unique.
Chapter 4

Consumer Theory I

4.1 Duality

4.1.1 Guiding Question (MWG 3.G.16)

Consider the expenditure function

\[ e(p, u) = \exp \left\{ \sum_l \alpha_l \log p_l + \left( \prod_l p_l^{\beta_l} \right) u \right\} \]  

(4.1)

a) What restrictions on \( \alpha_1, \ldots, \alpha_n, \beta_1, \ldots, \beta_n \) are necessary for this to be derivable from utility maximization?

b) Find the indirect utility that corresponds to it.

c) Verify Roy’s identity and the Slutsky equation

4.1.2 Background

So far, we have examined the consumer’s problem (CP) from two different perspectives. The expenditure minimization problem (EMP) chooses consumption to minimize the cost needed to achieve a target utility level. As such, its objective is linear, while its constraint set is a nonlinear convex set. The utility maximization problem (UMP) chooses consumption to maximize the utility that can be achieved without exceeding a budget. As such, its objective is nonlinear and concave and its constraint set is linear. These similarities are often described with the catch-all term “duality”. For a more precise presentation of this idea, in terms of mathematical objects called “support functions”, see the note on the duality theorem.

The two problems we have just discussed can be expressed mathematically as

\[ v(p, \omega) = \arg \max_x u(x) \]
\[ \text{s.t. } p \cdot x \leq \omega \]  

(4.2)
and
\[ e(p, \psi) = \arg \min_h p \cdot h \quad \text{s.t.} \quad u(h) \geq \psi \] (4.3)

But, note that we can also express them in terms of each other. We can eliminate the need for the actual consumption bundle to come into the optimization problem, using utility and expenditure directly as our choice variables. Mathematically, we reexpress the UMP and the EMP as
\[ v(p, \omega) = \arg \max_u u \quad \text{s.t.} \quad e(p, u) \leq \omega \] (4.4)
and
\[ e(p, \psi) = \arg \min_w w \quad \text{s.t.} \quad v(p, w) \geq \psi \] (4.5)
respectively. Let’s use these alternate formulations of the CP to derive some useful duality identities. Consider (4.4) with \( \omega = e(p, \psi) \). Because \( e(p, u) \) is increasing in \( u \), the biggest that we can make \( u \) is a value such that \( e(p, u) = e(p, \psi) \). So, we find that the maximum \( u \) value is \( \psi \), and hence \( v(p, e(p, \psi)) = \psi \).

Similarly, consider (4.5) with \( \psi = v(p, \omega) \). Since \( v(p, w) \) is increasing in \( w \), the smallest that we can make \( w \) is a value such that \( v(p, w) = v(p, \omega) \). So, we find that the minimum \( w \) value is \( \omega \), and hence \( e(p, v(p, \omega)) = \omega \).

So, using the properties of the value functions, we have related \( v \) to \( e \). How now might we relate \( h \) to \( x \)? Roy’s identity is derived from applying the envelope theorem to optimization problem (4.2).
\[ x(p, \omega) = - \frac{\nabla_p v}{\partial v/\partial w} \] (4.6)

If we apply the envelope theorem to (4.4), we find that \( \nabla_p v(p, \omega) = -\lambda \nabla_p e(p, u^*) \) and \( \frac{\partial w}{\partial w} = \lambda \) where \( \lambda \) is the appropriate multiplier. Hence, using the previous equation, we can write the Marshallian demand as
\[ x(p, \omega) = \nabla_p e(p, u^*) \] (4.7)
By definition, \( u^* = v(p, \omega) \), and Shepard’s lemma tells us that \( \nabla_p e \) is just the Hicksian demand. Hence, we have derived the identity
\[ x(p, \omega) = h(p, v(p, \omega)) \] (4.8)

Similarly, if we apply the envelope theorem to (4.3), we derive Shepard’s lemma
\[ h(p, \psi) = \nabla_p e(p, \psi) \] (4.9)
If we apply the envelope theorem to (4.5) though, we find \( \nabla_p e(p, \psi) = -\mu \nabla_p v(p, w^*) \), where \( \mu \) is the appropriate multiplier. By the definition of the optimization problem, \( w^* = e(p, \psi) \). But, what is \( \mu \). The first order condition for (4.5) tells us that
\[ 1 - \mu \frac{\partial w}{\partial w}(p, w^*) = 0 \] (4.10)
Hence, $\mu = \frac{1}{\partial v/\partial w}$, and we come to

$$h(p, \psi) = -\nabla_p v(p, w^*) = x(p, w^*)$$

(4.11)

where the last equality comes from Roy’s identity. Plugging in for $w^*$, we find that

$$h(p, \psi) = x(p, e(p, \psi))$$

(4.12)

The relations we have derived helps us to move from the EMP to the UMP and back. They will prove quite useful. Anytime you have already calculated the solution to one of these problems, you can use the duality identities we have just derived to get the solution to the other problem. In this way, these identities will often half the amount of work you have to do.

### 4.1.3 Solution to guiding question

a) To ensure that this expenditure is rational, we need

1) homogeneity of degree 1 in $p$
2) strictly increasing in $u$ and non-decreasing in $p_l$.
3) concave in $p$
4) continuous in $p$ and $u$

Which of these is satisfied straightaway? 4) and 2)$u$. So, let’s start with 2)$p_l$.

$$\frac{\partial e}{\partial p_k} = \frac{e(p, u)}{p_k} \left\{ \alpha_k + u/\beta_k \prod_l p_l^{\beta_l} \right\}$$

(4.13)

Where can we find problems with this? If we can construct any prices that make the right-hand side of the equation negative, then the expenditure cannot be rational. So, let’s look for weird prices that might do just that. If $\beta_k < 0$, then

For $\beta_l, \forall l$ send $p_l \to 0$ if $\beta_l < 0$
send $p_l \to \infty$ if $\beta_l > 0$

Note that these arrows are meant to convey the idea of sending the prices to an extreme in one direction or another, but not the idea of a formal limit. So, $p \to \infty$ just means to set $p$ sufficiently high. Our scheme is constructed to make the coefficient of $\beta_k$ large and positive. Following it, we can clearly make the negative $\beta_k$ term in the sum larger in magnitude than the $\alpha_k$ term. But this violates rationality of the expenditure. So, our assumption of a negative $\beta$ must not be allowed. Hence, $\beta_l \geq 0, \forall l$.

What if $\alpha_k < 0$? We know that $\beta_l \geq 0, \forall l$. If one of the $\beta_l$’s is strictly positive, then we can send its corresponding price close to zero, essentially
leaving a naked and negative $\alpha_k$. If $\beta_l = 0, \forall l$, then, $\alpha_k$ is already left alone and negative. So, $\alpha_k < 0$ violates rationality. Thus far, then, our conditions are

$$\beta_l, \alpha_l \geq 0, \forall l$$

(4.14)

OK, now let’s look at condition 1):

$$e(\lambda p, u) = \exp \left\{ \sum_l \alpha_l [\log p_l + \log \lambda] + \left[ \prod_l (\lambda p_l)^{\beta_l} \right] u \right\}$$

(4.15)

$$= \lambda^\sum \alpha_l \exp \left\{ \sum_l \alpha_l \log p_l + \lambda^\sum \beta_l \left( \prod_l p_l^{\beta_l} \right) u \right\}$$

(4.16)

So, to get homogeneity of degree 1, we need

$$\sum \alpha_l = 1$$

(4.17)

$$\sum \beta_l = 0$$

(4.18)

But $\beta_l \geq 0, \forall l$, so our final conditions are

$$\sum \alpha_l = 1$$

(4.19)

$$\beta_l = 0, \forall l$$

(4.20)

From here we can simplify $e(p, u)$ to

$$e(p, u) = e^u e^{\sum \alpha_l \log p_l^n} = e^u \prod p_l^{\alpha_l}$$

(4.21)

This is clearly increasing in $u$ (condition 4) and is concave in $p$ (condition 3) since $\alpha_l \in [0, 1], \forall l$. So, we have completely characterized the set of $\alpha_l$ and $\beta_l$ for which $e(p, u)$ is rationalizable:

$$\sum \alpha_l = 1 \quad \beta_l = 0, \forall l$$

(4.22)

b) Here, we see that expenditure and indirect utility are value functions for dual optimization problems. Hence, we would expect that they are closely related. We use the fundamental relation between the EMP and the UMP, $e(p, v(p, w)) = w$. Why is this so?

- Say $e(p, v(p, w)) > w$. Consider the arg max from the Marshallian\(^1\) problem used to derive $v(p, w)$. Call it $x^*$. By the constraint of the Marshallian problem, $p \cdot x^* \leq w$. Since $x^*$ is also the arg max of problem, it attains the maximum utility, $v(p, w)$ which is also, in our case, the target utility of the Hicksian\(^2\) problem. So we have constructed a solution with lower expenditure, contradicting optimality.

\(^1\) $v(p, w) = \max_x u(x) \text{ s.t. } p \cdot x \leq w$

\(^2\) $e(p, u) = \min_h p \cdot h \text{ s.t. } u(h) \geq u$
• Say \( e(p, v(p, w)) < w \). Then \( \exists x' \) such that \( p \cdot x' < w \) and \( u(x') \geq v(p, w) \). But since \( v(p, w) \) is a value function, we find that \( x' \) must be in the arg max of the Marshallian problem as well. But, if \( u(x) \) is locally nonsatiated, then this violates Walras’ Law, leading to a contradiction.

Thus, \( e(p, (v(p, w))) = w \). Now, using this relation, we can move forward.

\[
e(p, v(p, w)) = w = e^{v(p,w)} e^{\sum \log p_i^a_i} \tag{4.23}
\]

\[
e^v = we^{-\sum \log p_i^a_i} \tag{4.24}
\]

\[
v(p, w) = \log w - \sum \log p_i^a_i \tag{4.25}
\]

\[
e^{v(p,w)} = \frac{w}{p} \alpha_l \tag{4.26}
\]

c) Now, Roy’s identity says

\[
x_i = -\frac{\partial v/\partial p_i}{\partial v/\partial w} \tag{4.27}
\]

But we don’t have the Marshallian demand to plug in - we have the expenditure function. Can we use the fact that these problems are duals?

\[
h_i(p, u) = \frac{\partial e}{\partial p_i} = e(p, u) \frac{\alpha_i}{p_i} \tag{4.28}
\]

We know that \( u = v(p, w) \) relates the EMP to the UMP, so

\[
x_i(p, w) = e(p, v(p, w)) \frac{\alpha_i}{w} = \frac{w \alpha_i}{p_i} \tag{4.29}
\]

And we verify with Roy’s identity:

\[
x_i = +\frac{\alpha_i}{p_i} \frac{1}{w} = \frac{w \alpha_i}{p_i} \checkmark \tag{4.30}
\]

Now, what about Slutsky? The general form of this equation\(^4\) is

\[
\frac{\partial h_i(p, \bar{u})}{\partial p_k} = \frac{\partial x_i(p, \bar{w})}{\partial p_k} + \frac{\partial x_i(p, \bar{w})}{\partial w} x_k(p, \bar{w}) \tag{4.31}
\]

where \( \bar{w} = e(p, \bar{u}) \) (and similarly \( \bar{u} = v(p, \bar{w}) \)). The change in compensated (Hicksian) demand consists of a price effect and a wealth effect. Notice that the Slutsky equation only holds at set points - those that correspond to each

\(^3\)Take a moment to derive this. The Lagrangian of the Marshallian problem is \( \mathcal{L} = u(x) + \lambda(w - p \cdot x) \). \( \frac{\partial \mathcal{L}}{\partial p_i} = -\lambda x_i \) and \( \frac{\partial \mathcal{L}}{\partial w} = \lambda \). So, taking the ratio of the two, we can get \( x_i \). Finally, the envelope theorem tells us that derivatives of the Lagrangian (evaluated at the arg max) are equal to derivatives of the value function. So, \( x^* = -\frac{\partial v/\partial p_i} {\partial v/\partial w} \).

\(^4\)This equation is also easy to derive. Just take the derivative of \( h(p, v(p, w)) = x(p, w) \) which is just the fundamental dual relation expressed in terms of arg maxes instead of value functions. Then, use the identity \( \frac{\partial e}{\partial p_i} = h_i(p, u) \) which is easily attained from the envelope theorem.
other via the two versions of the consumer’s problem. So, if we want the Slutsky equation in terms of target utility \( u \), we can write

\[
\frac{\partial h_l(p, u)}{\partial p_k} = \frac{\partial x_l(p, e(p, u))}{\partial p_k} + \frac{\partial x_l(p, e(p, u))}{\partial w} x_k(p, e(p, u)) \tag{4.32}
\]

and if we want it in terms of wealth, we write

\[
\frac{\partial h_l(p, v(p, w))}{\partial p_k} = \frac{\partial x_l(p, w)}{\partial p_k} + \frac{\partial x_l(p, w)}{\partial w} x_k(p, w) \tag{4.33}
\]

Notice, that even in the equation that is supposedly in terms of \( u \), we have \( \partial/\partial w \). What does this mean? The standard notation is misleading here. It means, not a partial derivative with respect to \( w \), but a partial derivative with respect to the second argument of the function. Since \( x \) is usually given as a function of \((p, w)\), we refer to a derivative with respect to the second argument as a derivative with respect to \( w \). This is usually not misleading, but when we tack something that is not a \( w \) in the second argument, then things become confusing. Often to avoid this sort of confusion, economists refer to derivatives by number. So instead of writing \( \partial x/\partial w \), often we see \( \partial_x x \) (operator notation) or \( x_2 \) (subscript notation). All of these have their advantages and disadvantages. \( \partial x/\partial p \) is familiar and usually doesn’t cause trouble. Subscript notation can get confusing if \( x \) is already a vector and we need to refer to the derivative \( l^\text{th} \) component of \( x \) with respect to the second argument. For this reason, for maximum clarity, operator notation is sometimes convenient. In this notation, the Slutsky equation can be written in terms of target utility as

\[
\partial_{1k} h_l(p, u) = \frac{\partial x_l(p, e(p, u))}{\partial p_k} + \frac{\partial x_l(p, e(p, u))}{\partial w} x_k(p, e(p, u)) \tag{4.34}
\]

In short, partial derivatives are never taken with respect to a variable. Only full derivatives are actually taken with respect to a variable. Partial derivatives are taken with respect to an argument of a function\(^5\).

Back to the problem, we can calculate the left-hand side of (4.31) from (4.28)\(^6\)

\[
\frac{\partial h_l}{\partial p_k} = \frac{\partial e}{\partial p_k} \frac{\alpha_l}{p_l} - e(p, u) \frac{\alpha_l}{p_l^2} \delta_{l,k} \tag{4.35}
\]

\[
= h_k \frac{\alpha_l}{p_l} - e(p, u) \frac{\alpha_l}{p_l^2} \delta_{l,k} \tag{4.36}
\]

\[
= e(p, u) \left\{ \frac{\alpha_l \alpha_l}{p_k p_l} - \frac{\alpha_l}{p_l^2} \delta_{l,k} \right\} \tag{4.37}
\]

\[
= \frac{w}{\text{by Walras’ Law}} \left\{ \frac{\alpha_k \alpha_l}{p_k p_l} - \frac{\alpha_l}{p_l^2} \delta_{l,k} \right\} \tag{4.38}
\]

\(^5\)Still confused? Consider \( f(x, y) = xy \). \( \frac{\partial f}{\partial y}(x, y) = x \). But, also \( \frac{\partial f}{\partial y}(x, x) = x = \partial_2 f(x, x) \)

\(^6\)The Kronecker delta is defined as \( \delta_{l,k} = \begin{cases} 1 & l = k \\ 0 & l \neq k \end{cases} \)
Calculating the right-hand side of (4.31) from (4.30):

\[
\frac{\partial x_l}{\partial p_k} = -\frac{w\alpha_l}{p_l^2} \delta_{l,k} \tag{4.39}
\]

\[
\frac{\partial x_l}{\partial w} = \frac{\alpha_l}{p_l} \tag{4.40}
\]

\[
\frac{\partial x_l}{\partial p_k} + \frac{\partial x_l}{\partial w} x_k = -\frac{w\alpha_l}{p_l^2} \delta_{l,k} + \frac{\alpha_l}{p_l} \frac{w\alpha_k}{p_k} + \frac{\alpha_l}{p_l} \frac{w\alpha_k}{p_k} \tag{4.41}
\]

\[
= w \left\{ \frac{\alpha_l \alpha_k}{p_l p_k} - \frac{\alpha_l}{p_l^2} \delta_{l,k} \right\} \tag{4.42}
\]

(4.38) and (4.42) match, so we have verified that the Slutsky equation holds.

\section*{4.2 Homotheticity}

\subsection*{4.2.1 Guiding Question}

Suppose Consumer D has rational, continuous, and locally non-satiated preferences, and that her preferences are also homothetic, i.e. if \( x \sim x' \) then \( \alpha x \sim \alpha x' \). What can you conclude about how Consumer D’s Marshallian demand \( x(p, w) \) will vary with \( w \)?

\subsection*{4.2.2 Background}

A helpful characterization of homotheticity is that indifference curves dilated about the origin are also indifference curves. This follows immediately from the definition given in the problem. Homotheticity is usually presented in terms of preferences, but this definition in terms of the geometry of indifference curves will almost always be more useful in solving problems.

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{dilation.png}
\caption{Illustration of dilation property of homothetic indifference curves.}
\end{figure}
4.2.3 Solution to guiding problem

Homotheticity implies that all indifference curves are simply dilations of each other about the origin. Simple geometry then tells us that a ray from the origin will cross every indifference curve at the same angle. So say we have the optimizer for $(p, w)$. Call it $x^*$. The plane defined by $p \cdot x = w$ is crossed by the ray from the origin through $x^*$ at the same angle it crosses the indifference curve, because of the tangency condition at the optimum. If we increase $w$ to $\alpha w$, we have a new plane parallel to the old one that crosses our ray at $\alpha x^*$ at the same angle. But the ray crosses all indifference curves at the same angle as well. So this new plane is tangent to the new indifference curve. Hence $\alpha x^*$ is the optimum for wealth level $\alpha w$. So Marshallian demand will vary directly with wealth.

4.3 Aggregation and the representative consumer

4.3.1 Guiding question (MWG 4.B.2)

Suppose that there are $I$ consumers and $L$ commodities. Consumers differ only by their wealth levels $w_i$ and by a taste parameter $s_i$, which we could call family size. Thus, denote the indirect utility function of consumer $i$ by $v(p, w_i, s_i)$. The corresponding Walrasian demand function for consumer $i$ is $x(p, w_i, s_i)$.

a) Fix $(s_1, \ldots, s_I)$. Show that if for any $(w_1, \ldots, w_I)$, aggregate demand can be written as a function of only $p$ and aggregate wealth $w = \sum_i w_i$ (or, equivalently, average wealth), and if every consumer’s preference relationship $\succ_i$ is homothetic, then all these preferences must be identical [and so $x(p, w_i, s_i)$ must be independent of $s_i$].

b) Give a sufficient condition for aggregate demand to depend only on aggregate wealth $w$ and $\sum_i s_i$ (or, equivalently, average wealth and average family size).

4.3.2 Background

Thus far, we have mainly focused on individual behavior. What about the aggregate behavior of society as a whole? How do we aggregate our data? Let’s start by defining aggregate demand

$$x(p, w) \equiv \sum_{i=1}^{I} x_i(p, w_i)$$

(4.43)

So a good question to ask, for starters, is when is it the case that aggregate demand only depends on $p$ and aggregate wealth, $\sum w_i \equiv W$? The analytical idea here is to be able to treat society as rationally equivalent to one representative agent who possesses all the wealth in society. Subtly, what this means is that the demand of society as a whole does not depend on the wealth distribution.
Whoa! Isn’t this a stupid thing to try and prove? It is clearly not true in the real world. Why would we even attempt such a proof? The answer is that it is a general trend in economic thought to prove something ridiculous and then to analyze the assumptions that were needed to get there. In this way, we can see what went wrong in the first place. By doing so, we are able to determine what is at the heart of the more realistic and more complicated behavior that we observe.

For example, consider the Coase Theorem. It says that if you assign clear property rights, even in the presence of an externality, and efficient outcome will ensure, granted there are no transaction costs. This ridiculous conclusion lets us know that transaction costs must be an important part of a model that seeks to explain the real world. As another example consider the Modigliani-Miller theorem, which says that in the absence of taxes, bankruptcy costs, and asymmetric information, and in an efficient market, the value of a firm is unaffected by how that firm is financed. Again, the conclusion is clearly false, but it sheds light on what is important to consider. Anyways, back to math.

Let’s start with a wealth distribution

\[
 w = (w_1, w_2, \ldots, w_I) \quad (4.44)
\]

Now, consider an infinitesimal redistribution, \( dw \) leading to the new distribution

\[
 w' = (w_1 + dw_1, w_2 + dw_2, \ldots, w_I + dw_I) \quad (4.45)
\]

To make it a pure redistribution, what constraint must we impose?

\[
 \sum dw_i = 0 \quad (4.46)
\]

And, how will aggregate demand change as a result of the redistribution. Using the chain rule,

\[
 dx(p, w, dw) = \sum_i \frac{\partial x_i}{\partial w} dw_i \quad (4.47)
\]

For society to have its aggregate demand independent of distribution, we need

\[
 dx = 0 \quad (4.48)
\]

If \( dw_i \) is arbitrary, subject to the constraint, then \( dx = 0 \) implies

\[
 \frac{\partial x_i}{\partial w} = \frac{\partial x_j}{\partial w}, \quad \forall i, j \quad (4.49)
\]

A necessary and sufficient condition for this to hold is for the individual indirect utilities to be of Gorman form:

\[
 v_i(p, w_i) = a_i(p) + b(p)w_i \quad (4.50)
\]

This form works because it uses quasi-linear separation to keep individual wealth levels away from the other idiosyncrasies of individual demand. You will show sufficiency in the homework; necessity is quite difficult \(^7\) and not advised (unless you aren’t getting enough work).

\(^7\)The original paper is W. M. Gorman. On a class of preference fields, Metroeconomica, 13, August 1961, 53-56.
4.3.3 Solution to guiding question

a) With zero wealth, assuming positive prices, the consumer cannot afford to purchase anything. Hence,
\[ x_i(p, 0, s_i) = 0 \quad (4.51) \]

Which wealth distributions might be instructive to consider?

Give all wealth to agent \( i \) in one distribution and to agent \( j \) in another. What is the expression for aggregate demand in these two cases? If consumer \( i \) has all the wealth, then
\[ X(p, 0, \ldots, 0, W, 0, \ldots) = x(p, W, s_i) \quad (4.52) \]

But, this must hold for all \( i \), so we conclude
\[ x(p, W, s_i) = x(p, W, s_j), \forall i, j \quad (4.53) \]

But if this holds for wealth level \( W \), then it must hold for all levels of wealth, by homotheticity of preferences (since we can just say that \( x(p, \alpha W, s_i) = \alpha x(p, W, s_i) \)). So
\[ x(p, w, s_1) = x(p, w, s_j), \forall i, j \quad (4.54) \]

or, more succinctly, that individual preferences are independent of \( s \) (i.e. identical).

b) So perhaps our condition should have to do with the form of the indirect utility function, like the Gorman condition. Any ideas?

The Gorman form was able to aggregate over wealth by using quasi-linear separation. What if we did the same sort of thing with \( s \)? Consider
\[ v_i(p, w_i, s_i) = a_i(p) + b(p)w_i + c(p)s_i \quad (4.55) \]

Does this work? How could we check? Let’s run Roy’s identity up the flagpole and see who salutes.
\[ x_i = -\frac{\nabla_p v_i}{\partial v_i/\partial w} = -\frac{[\nabla a_i(p) + \nabla b(p)w_i + \nabla c(p)s_i]}{b(p)} \quad (4.56) \]

And aggregating, we find
\[ \sum x_i = -\frac{1}{b(p)} \sum \nabla a_i(p) - \frac{\nabla b(p)}{b(p)} \sum w_i - \frac{\nabla c(p)}{b(p)} \sum s_i \quad (4.57) \]

So, our aggregate demand depends only on \( W \) and \( \sum s_i \); we have shown the existence of a representative agent.

This is a simple example of a representative agent theorem. These crop up is almost all fields of economics and generally have the same proof strategy.

---

\(^8\) X is the aggregate demand as a function of the wealth distribution and the price.
Chapter 5

Consumer Theory II

5.1 Welfare measures and path dependence

5.1.1 Guiding Question (2005 202N Midterm, Question 5)

Suppose that $X = \mathbb{R} \times \mathbb{R}_+^{n-1}$, and that goods 2 through $n$ can be put in categories 2 through $m$ such that

$$u(x) = x_1 + \varphi_2(z_2) + \varphi_3(z_3) + \cdots + \varphi_m(z_m) \quad (5.1)$$

where $z_i$ is a vector of the goods in category $i$. Assume that $p_1 = 1$ is fixed throughout the problem, and let $q_i$ be a vector of the prices of the goods in category $i$.

a) Show that Marshallian demand for the goods in category $i$ is the solution to the maximization problem

$$\max_{z_i \geq 0} \left\{ \varphi_i(z_i) - q_i \cdot z_i \right\} \quad (5.2)$$

and therefore does not depend on wealth or on the prices of goods in other categories.

b) Show that Hicksian demand for the goods in category $i$ similarly does not depend on the target utility or the prices of goods outside of category $i$.

c) Consider a price change where only the prices of goods in category $i$ change. Show that Equivalent Variation and Compensating Variation are equal to each other and do not depend on the prices of goods outside category $i$ or the initial or subsequent wealth or utility level.

(If you get stuck, you can get partial credit for showing this is true when only one price changes.)

**Extra Credit.** Consider a price change from $p$ to $p' = (1, p'_2, p'_3, \ldots, p'_n)$. We can rewrite the expression for Compensating Variation as

$$CV = e(p, u) - e(p', u) = e((1, p_2, p_3, \ldots, p_{n-1}, p_n), u) - e((1, p'_2, p'_3, \ldots, p'_{n-1}, p'_n), u) + e((1, p'_2, p'_3, \ldots, p'_{n-1}, p_n), u) - e((1, p'_2, p'_3, \ldots, p'_{n-1}, p'_n), u) + \cdots + e((1, p'_2, p'_3, \ldots, p'_{n-1}, p_n), u) - e((1, p'_2, p'_3, \ldots, p'_{n-1}, p'_n), u)$$

53
It’s tempting to interpret each line of the equation above as the welfare effect of each individual price change. However, with typical utility functions, we can’t do this, as our calculations would depend on the order in which we made the price changes.

Show that in this case (with the utility function given above), we can sensibly calculate a separate CV for the price changes of goods in each category, which are well-defined (do not depend on how the calculation if performed) and will add up to the total CV.

### 5.1.2 The easy part (a and b)

a) Writing down the consumer’s problem (CP):

\[
v(q, w) = \max_x x_1 + \varphi_2(z_2) + \cdots + \varphi_m(z_m)
\]

\[
s.t. \quad x_1 + \sum_{i=2}^{m} q_i \cdot z_i \leq w
\]

(5.3)

How can we proceed now, bearing in mind that we know nothing about the \(\varphi_i\)’s?

Note that the constraint will bind due to Walras’ Law. We don’t know about the \(\varphi_i\)’s, but \(u(x)\) is locally nonsatiated since it is always increasing in \(x_i\). This means we can simplify the CP by plugging in the binding constraint:

\[
v(q, w) = \max_x w - \sum_{i=2}^{m} q_i \cdot z_i + \varphi_2(z_2) + \cdots + \varphi_m(z_m)
\]

(5.4)

How is this further simplified? Note that the only \(w\) in the equation passes right through the max operator.

\[
v(q, w) = w + \sum_{i=2}^{m} \max_{z_i} \{\varphi_i(z_i) - q_i \cdot z_i\}
\]

(5.5)

So, we have shown that Marshallian demand does not depend on wealth or the prices of items in other groups.

Also, note that if all demands besides \(x_1\) are set independent of wealth, then if wealth \(w\) is small enough, \(x_1\) will go negative. This is a property of problems with quasilinear preferences, and usually the assumption that the linear good is non-negative is relaxed, as it is in this problem, since the choice set is \(\mathbb{R} \times \mathbb{R}_+^{n-1}\).

b) Does the same approach work here? Yes, the no excess utility property leads to the constraint binding - everything else follows as before. But, let’s try an alternate approach, for the sake of learning. How might we solve b) using the duality of the Hicksian and Marshallian demands?

With duality, we can relate expenditure and indirect utility. If \(w = e(p, u)\), then

\[
v(p, e(p, u)) = e(p, u) + \sum_{i=2}^{m} \max_{z_i} \{\varphi_i(z_i) - q_i \cdot z_i\}
\]

(5.6)
To simplify, define the partial indirect utility by
\[
\psi_i(q_i) \equiv \max_{z_i \geq 0} \{ \varphi_i(z_i) - q_i \cdot z_i \} \tag{5.7}
\]

Then, we have
\[
v(p, e(p, u)) = e(p, u) + \sum_{i=2}^{m} \psi_i(q_i) \tag{5.8}
\]

Now, noticing that \( v(p, e(p, u)) = u \), we have come to an expression for the expenditure:
\[
e(p, u) = u - \sum_{i=2}^{m} \psi_i(q_i) \tag{5.9}
\]

Now, we can use Shepard’s Lemma (the envelope theorem) to pull out Hicksian demand:
\[
h = \nabla_p e \tag{5.10}
\]
\[
h_i(p, u) = \frac{\partial e(p, u)}{\partial p_i} = - \frac{\partial \psi_i(q_i)}{\partial p_i} \tag{5.11}
\]

So, Hicksian demand is independent of target utility and outside prices.

5.1.3 Background on welfare measures

Let’s quickly skim the ideas behind the equivalent and compensating variations (EV and CV) as welfare measures. Consider a price change from \( p \) to \( p' \). How might we measure the change in welfare?

The initial reaction for most people is to use \( v(p, w) \), holding wealth constant. Hence
\[
v(p, w) - v(p', w) \tag{5.12}
\]
is how much more utility a consumer can afford at \( p \) than at \( p' \), holding wealth constant. Is there any problem with this?

Utility is an ordinal concept. The positivity or negativity of our quantity yields a preference, but it tells us nothing about how much better or worse off our agent is. What can we do about this dilemma?

We can use expenditures since they are measured in “real” units (dollars, euros, yuan, etc.). So what about
\[
e(p, v(p, w)) - e(p', v(p', w)) \tag{5.13}
\]
as a utility measure? Is there a problem here?

Duality ensures us that this measure is always equal to zero. We are measuring the change twice - once in our target utility and again in the expenditure. This is the problem - changing utility as well as price is the wrong idea. If we are comparing expenditures, then they need to be expenditures relative to the same utility level.
This measures welfare because if some utility $u$ costs more under $p'$ than under $p$, then consumers are clearly worse off. Using this philosophy of measurement, we have two possible welfare measures. Assuming constant wealth $w$, we let $u$ be the value of the indirect utility under prices $p$ and $u'$ be the value under prices $p'$. Then the **compensating variation** is given by

$$CV = e(p, u) - e(p', u) = e(p, v(p, w)) - e(p', v(p, w)) = w - e(p', v(p, w))$$ (5.14)

and the **equivalent variation** is given by

$$EV = e(p, u') - e(p', u') = e(p, v(p', w)) - e(p', v(p', w)) = e(p, v(p', w)) - w$$ (5.15)

The $CV$ represents how much expenditure the consumer must give up under $p'$ to maintain the same utility she had under $p$. Similarly, the $EV$ represents how much extra expenditure the consumer must be allowed under $p$ to attain the same utility level she had attained under $p'$.

As a semantic note, there isn’t much logic behind the names given to these two very similar concepts. I suggest the mnemonic that C is before E, alphabetically, so $CV$ is measured at $u$ while $EV$ is measured at $u'$.

### 5.1.4 The harder part (c)

Recall our expression for the expenditure function in terms of the partial indirect utilities in (5.9). The two welfare measures are then given by

$$CV = -\sum_{i=2}^{m} [\psi_i(q_i) - \psi_i(q_i')]$$ (5.16)

$$EV = -\sum_{i=2}^{m} [\psi_i(q_i) - \psi_i(q_i')]$$ (5.17)

Since $u$ enters $e(p, u)$ linearly, $CV$ and $EV$ simply lose any mention of $u$, and since Hicksian demand was independent of wealth and outside prices, we could contruct the $\psi_i$’s that made this possible. So $CV = EV$ and both are independent of $w$. What’s more, this isn’t only true when the price changes are restricted to one category. $CV = EV$ for arbitrary price changes. And, if price changes are limited to one category, then only one of the elements in the sums above is non-zero, so we find that the $CV$ and the $EV$ for a price change that only affects one category are independent of prices outside of that category.

### 5.1.5 Background on path dependence and conservative fields

Before we attempt the extra credit, let’s examine integral welfare measures and the idea of path independence.

If we move one price $p_i$ to $p'_i$, it is pretty clear that we can express the $CV$ as

$$CV = e(p, u) - e(p', u) = \int_{p_i}^{p'_i} \frac{\partial e}{\partial p_i}(p, u) dp_i$$ (5.18)
5.1. WELFARE MEASURES AND PATH DEPENDENCE

in a style reminiscent of our derivation of the integral form of the envelope theorem. Shepard’s lemma then tells us that this reduces to

\[ CV = \int_{p_i}^{p_i'} h_i(p, u) dp_i \]  
\[ EV = \int_{p_i}^{p_i'} h_i(p, u') dp_i \]  

But what about cases with multiple price changes? What if we have \((p_i, p_j) \rightarrow (p_i', p_j')\)? Does it matter if we first send \(p_i \rightarrow p_i'\) and then \(p_j \rightarrow p_j'\) or if we do it the other way around \((p_j \rightarrow p_j'\) and then \(p_i \rightarrow p_i'\))? To ask a more general question, does the path we follow in price space from \(p\) to \(p'\) matter in the evaluation of the integral welfare measure? Let’s diverge to vector calculus for a second. Say \(C_1\) and \(C_2\) are two curves in price space that have the same endpoints. Then, if \(F = \nabla f\) for some scalar function \(f\), we have

\[ \int_{C_1} F \cdot dp = \int_{C_2} F \cdot dp \]  

or alternatively, for any closed path,

\[ \oint F \cdot dp = 0 \]

But, by Shepard’s Lemma, we see that if we use \(e(p, u)\) as the scalar function \(f\) (our **potential**), then the corresponding \(F\) (our **field**) is the Hicksian demand vector \((\nabla_p e(p, u) = h(p, u))\).

So, we have path independence in price space (**subject to the differentiability of \(e\)**). Thus, there is no ambiguity in defining our welfare measures by

\[ CV = \int_{p}^{p'} h(p, u) \cdot dp \]
\[ EV = \int_{p'}^{p} h(p, u') \cdot dp \] (5.24)

since the path from \( p \) to \( p' \) used the evaluate the integrals is inconsequential. Thus the path integral formulation matches to the simpler formulation of expressing \( CV \) and \( EV \) as differences of some scalar state function, the expenditure.

Path independence of a welfare measure between two endpoints in price space gives meaning to the welfare measure. Without path independence, we cannot unambiguously say anything about the welfare change without specifying a price path. As an illustration, consider the utility function \( \sqrt{xy} \). Solving for expenditure, we find

\[
\begin{align*}
\text{min}_{x,y} \quad & p_x x + p_y y \\
\text{s.t.} \quad & \sqrt{xy} \geq u
\end{align*}
\]

\( L = p_x x + p_y y - \lambda (\sqrt{xy} - u) \) (5.26)

First-order conditions

\[
\begin{align*}
[x] : & \quad p_x = \frac{1}{2} \lambda \frac{x}{\sqrt{y}} \\
[y] : & \quad p_y = \frac{1}{2} \lambda \frac{y}{\sqrt{x}} \\
[\lambda] : & \quad \sqrt{xy} = u \Rightarrow y = \frac{u^2}{x}
\end{align*}
\]

Combining these two conditions yields

\[
\begin{align*}
\frac{x p_x}{p_y} &= \frac{u^2}{x} \quad (5.28) \\
x^* &= u \sqrt{\frac{p_y}{p_x}} \quad (5.29) \\
y^* &= u \sqrt{\frac{p_x}{p_y}} \quad (5.30) \\
e(p, u) &= 2u \sqrt{p_x p_y} \quad (5.31)
\end{align*}
\]

Now consider \( p = (1, 1) \) and \( p' = (4, 4) \) given initial \( w = 2 \), so that \( u = 1 \). Calculating \( CV \)

\[
CV = e(p, u) - e(p', u) = \frac{e((1, 1), 1)}{2} - \frac{e((4, 4), 1)}{8} = -6 \quad (5.32)
\]

Now, let’s decompose this price change

\[
CV = \frac{e((1, 1), 1)}{2} - \frac{e((4, 1), 1)}{4} + \frac{e((4, 1), 1)}{4} - \frac{e((4, 4), 1)}{8} \quad (5.33)
\]

So, what is wrong with saying that the \(-2\) corresponds to the change in \( p_x \)? The whole problem is symmetric. We could just as easily have changed \( p_y \) first and then we would have associated the \(-2\) with it. \( x \) and \( y \) mix in our utility function, leading to mixing in our potential, \( e(p, u) \).

So, when can we reasonably split up our welfare measure into parts that correspond to different prices changes? The key to answering this is the “mixing” we
5.2. NESTED OPTIMIZATION AND THE LE CHÂTELIER PRINCIPLE

found in the example above. Let’s consider a path \( C \) that is parameterized by \( t \) such that

\[
C = \{ p(t) | t \in [0, \tau] \} \tag{5.34}
\]

where \( p \) is a price vector. Then, the path integral can be expressed

\[
CV(p(0), p(\tau), u) = \int_C h(p(t), u) \cdot dp(t) = \sum_i \int_0^\tau h_i(p(t), u)p_i'(t)dt \tag{5.35}
\]

where each element of the sum is the “component” of the \( CV \) due to \( p_i \). Path invariance is made manifest in the fact that \( CV \) only depends on the endpoints of \( C \), \( p(0) \) and \( p(\tau) \), and not the entire path. Now consider a set of indices, \( \sigma \), that correspond to prices, \( p_{\sigma} \). Then, we can define the “component” of the \( CV \) that corresponds to those prices as

\[
CV_\sigma(p(\cdot), u) = \sum_{i \in \sigma} \int_0^\tau h_i(p_\sigma(t), p_{\sigma}(t), u)p_i'(t)dt = \int_C h_\sigma \cdot dp_\sigma \tag{5.36}
\]

Since we are not sure about path independence, we express the \( CV_\sigma \) as a functional of the full path \( p(\cdot) \). So, when is this measure path independent? The vector calculus discussed above tells us that this is the case when we can express the integral as \( \int_{C_\sigma} \nabla_{p_\sigma} f \cdot dp_\sigma \) for some scalar function \( f \) (where \( C_\sigma \) is a path where \( p_{\sigma} \) does not move). So what is holding us back is the fact that our general path \( C \) moves \( p_\sigma \) and \( p_{\sigma} \). How can we dodge this problem?

If \( h_\sigma \) does not depend on \( p_{\sigma} \), then it doesn’t matter what \( C \) does with \( p_{\sigma} \), since it never appears in the integrand. So if we can derive a Hicksian demand \( h_\sigma \) that doesn’t depend on \( p_{\sigma} \) from some scalar expenditure with Shephard’s lemma \( (\nabla_{p_\sigma} e(p, u) = h_\sigma(p_\sigma, u)) \), then we have a path integral welfare measure that only depends on the endpoints in \( p_\sigma \).

This path invariance is what makes a welfare measure meaningful. Thus, whenever we can derive a \( h_\sigma \) from the expenditure that only depends on the prices \( p_\sigma \), then we can meaningfully split the welfare difference from a change in price in a component due to changing \( p_\sigma \) and a component due to changing \( p_{\sigma} \).

5.1.6 The hardest part (Extra Credit)

Let \( \sigma \) represent all prices in a category. It is clear that \( h_\sigma \) is independent of \( p_{\sigma} \) for all categories, hence it makes sense to split up welfare change into components corresponding to each category.

5.2 Nested Optimization and the Le Châtelier Principle

5.2.1 Guiding Question (June 2004 Comp, Question 1)

We study a consumer problem with three goods, \( x, y, \) and \( z \), a budget of \( B \), and a price vector \( p = (p_x, p_y, p_z) \).
CHAPTER 5. CONSUMER THEORY II

a) Let \( u(x, y, z) = (xy)^{1/2} + z \). Let \( b \) denote the consumer’s combined spending on goods \( x \) and \( y \). Show that there is always an optimum at which \( b \) is equal to 0 or \( B \). Obtain expressions for the consumer’s indirect utility function and expenditure function.

b) Let \( u(x, y, z) = (xy)^{\alpha} + z \) where \( 0 < \alpha < 1/2 \). Assume that the optimal solution is interior and let \( b \) be the combined spending on goods \( x \) and \( y \). Derive a formula for \( b \) in terms of \( p \) and \( B \).

c) Let \( u(x, y, z) = x^\alpha y^\beta + w(z) \), where the two parameters are positive: \( \alpha, \beta > 0 \). Suppose that \( w \) is concave. What additional assumptions, if any, are required to conclude that combined spending \( b \) on goods \( x \) and \( y \) rises (at least weakly) when the consumer’s budget \( B \) rises? To conclude that goods \( x \) and \( y \) are normal?

For the remaining two parts, assume that good \( y \) is fixed in the short-run (for example, it may represent expenditures on housing).

d) Assume that \( u(x, y, z) = v(x, y) + z \), that \( v_{xy}(x, y) \equiv \frac{\partial^2 v}{\partial x \partial y} > 0 \), and that all optima are interior optima. If the government introduces a tax \( t > 0 \) to raise the price of good \( x \) to \( p_x + t \) per unit, what will happen to consumption of goods \( x \) and \( y \) and to the tax revenue in the short-run and in the long-run?

e) How does your answer to iv) change if you change the conditions so that \( v_{xy}(x, y) < 0 \)?

5.2.2 Background

There isn’t much background to be introduced here. Sometimes it helps to nest maximization problems (this is the trick to many comp questions), so we will practice by doing the problem above. Also, the Le Châtelier principle has a close tie with supermodularity, as will be demonstrated in the last two parts of the problem.

5.2.3 Solution to guiding question

a) How should we begin this problem?

Kuhn-Tucker is one approach. Give it a try. In theory, straight Kuhn-Tucker should work out, but for me, it quickly got confusing. Instead, let’s use a trick that takes advantage of the additive separability of our problem. Any ideas?

If we give a budget \( b \in [0, B] \) to the purchase of \( x \) and \( y \), we solve the following problem

\[
v(b) = \max_{x, y \geq 0} \sqrt{xy} \quad \text{s.t.} \quad p_x x + p_y y \leq b \quad (5.37)
\]

Then, our Lagrangian is given by

\[
\mathcal{L} = \sqrt{xy} + \lambda [b - p_x x - p_y y] \quad (5.38)
\]
What about positivity multipliers?

Our objective is clearly locally non-satiated, so the constraint will bind by Walras’ Law. If \( x = 0 \) or \( y = 0 \), then we arrive at a zero-valued objective. It is pretty clear that this won’t be the case at the optimum (unless \( b = 0 \)). Now for our first-order conditions:

\[
\begin{align*}
[x]: & \quad \frac{1}{2} \sqrt{\frac{y}{x}} = \lambda p_x \\
[y]: & \quad \frac{1}{2} \sqrt{\frac{x}{y}} = \lambda p_y
\end{align*}
\]

The constraint binds, so

\[
\begin{align*}
b = p_x x + p_y y = 2p_x x \Rightarrow \\
& \begin{cases}
x^* = \frac{b}{2p_x} \\
y^* = \frac{b}{2p_y}
\end{cases}
\end{align*}
\]

and we arrive at the value function

\[
v(b) = \frac{1}{2} b (p_x p_y)^{-1/2}
\]

Can we phrase the original optimization in terms of \( v(b) \)? Yes:

\[
\begin{align*}
\max_{b, z \geq 0} & \quad v(b) + z \\
s.t. & \quad b + z \leq B
\end{align*}
\]

Note we have normalized \( p_z \) to 1. Again, we have local non-satiation, so \( z = B - b \), and we are left with

\[
\max_{0 \leq b \leq B} v(b) + B - b
\]

So, our Lagrangian becomes

\[
\mathcal{L} = v(b) + B - b + \mu b + \nu(B - b)
\]

where \( \mu \) is \( b \)'s positivity multiplier and \( \nu \) is the multiplier for \( b \leq B \). First-order conditions then yield

\[
v'(b) = \frac{1}{2} (p_x p_y)^{-1/2} = 1 + \nu - \mu
\]

Positivity tells us that \( \mu, \nu \geq 0 \) and complementary slackness says

\[
\mu b = 0 \quad \nu(B - b) = 0
\]

What does this first-order condition tell us?

The multipliers \( \mu \) and \( \nu \) cannot both be strictly positive, since they represent contradictory constraints. If \( \sqrt{p_x p_y} = 1/2 \), then \( b \) can be any element of \([0, B]\)
(simply set $\mu = \nu = 0$).\footnote{While, this could be the case, the prices are exogenous parameters, so we cannot set them to be this way. This is an example of an unintended complication in a comp question. These happen quite frequently. The only guide I can give you is to consider what the testmakers intended to ask and not what they actually ask. Don’t get hung up on little details that come out in the problems. The profs do make mistakes in the comp problems, so be aware.} If we are not in this razor’s edge case, then, since $\mu, \nu \geq 0$ with only one possibly holding strictly, we have

$$\text{If } \sqrt{p_x p_y} > \frac{1}{2}, \text{ then } \mu > 0 \Rightarrow b = 0$$
$$\text{< } \frac{1}{2}, \text{ then } \nu > 0 \Rightarrow b = B$$ \hfill (5.47)

So regardless of prices, $\exists$ an optimum with either $b = 0$ or $b = B$.\footnote{See how the problem skirts the edge of being misstated. But it says to show that a $b = 0$ or a $b = B$ optimum always exists, not that it is the unique optimum. Comp questions often play these semantic games. Don’t be fooled!} Which of these is the solution depends upon the geometric mean of the prices. Plugging in for indirect utility is trivial.

$$v(p, B) = B \max \left( \frac{1}{2 \sqrt{p_x p_y}}, 1 \right)$$ \hfill (5.48)

Then, using duality, we have

$$v(p, e(p, u)) = u = e(p, u) \max \left( \frac{1}{2 \sqrt{p_x p_y}}, 1 \right)$$ \hfill (5.49)

So,

$$e(p, u) = \frac{u}{\max \left( \frac{1}{2 \sqrt{p_x p_y}}, 1 \right)}$$ \hfill (5.50)

b) We use the exact same idea here

$$v(b) = \max_{x, y \geq 0} \ (xy)^\alpha \ \text{ s.t. } p_x x + p_y y \leq b \ \alpha \in (0, \frac{1}{2})$$ \hfill (5.51)

$$\begin{cases}
[x] : \alpha x^{\alpha - 1} y = p_x \\
[y] : \beta y^{\alpha - 1} = p_y
\end{cases} \Rightarrow
\begin{align*}
x^* &= \frac{b}{2 p_x} \\
y^* &= \frac{b}{2 p_y}
\end{align*}$$ \hfill (5.52)

$$v(b) = \left( \frac{b^2}{4 p_x p_y} \right)^\alpha$$ \hfill (5.53)

Now, (normalizing $p_z$ again)

$$\max_{0 \leq b \leq B} v(b) + B - b$$ \hfill (5.54)

$$\mathcal{L} = v(b) + B - b + \mu b + \nu (B - b)$$ \hfill (5.55)

$$[b] : v'(b) = 1 + \nu p^0 (\text{assume interiority!})$$ \hfill (5.56)
Plugging in for \( v'(b) \):
\[
2\alpha b^{2\alpha - 1}(4p_xp_y)^{-\alpha} = 1
\]
(5.57)
\[
b^* = \left[ \frac{(4p_xp_y)^{\alpha}}{2\alpha} \right]^{1/(2\alpha - 1)}
\]
(5.58)
which is independent of \( B \)? How can this be? We were told to assume interiority, which is a terrible assumption.

Finally, note that a characteristic of the Cobb-Douglas utility is that spending on each good is a constant fraction of wealth. So, save time and just write this down. If \( u(x, y) = x^\alpha y^\beta \), then we will have \( p_1x_1 = \frac{\alpha}{\alpha + \beta}w \) and \( p_2x_2 = \frac{\beta}{\alpha + \beta}w \). Learning little tricks saves precious time on a comp, when time matters.

c) How do we ensure that \( b^* \) increases weakly with \( B \)? This is a job for Topkis, so we need to check for supermodularity in \((b, B)\). What is the objective? Let
\[
v(b) = \max_{x,y \geq 0} x^\alpha y^\beta \quad \text{s.t.} \quad p_x x + p_y y \leq b
\]
(5.59)
Then, the problem becomes
\[
\max_{0 \leq b \leq B} v(b) + w(B - b) \quad \equiv \varphi(B, b)
\]
(5.60)
Testing for supermodularity:
\[
\frac{\partial \varphi}{\partial B} = w'(B - b)
\]
(5.61)
\[
\frac{\partial^2 \varphi}{\partial B \partial b} = -w''(B - b)
\]
(5.62)
w is concave, thus we have
\[
\frac{\partial^2 \varphi}{\partial B \partial b} \geq 0
\]
(5.63)
Our objective is supermodular in \((B, b)\), hence \( b^*(B) \) is weakly increasing with \emph{no added assumptions}.

What about \( x \) and \( y \) being normal goods? What is a normal good? \emph{A good is normal if Marshallian demand increases with wealth.} Again, our nesting approach will help a lot. How?

If \( b^*(B) \) is weakly increasing, then we only have to look at the Marshallian demands from the nested problem, that is \( x^*(b) \) and \( y^*(b) \). The complete Marshallian demands are thus \( x^*[b^*(B)] \) and \( y^*[b^*(B)] \). By the chain rule,
\[
\frac{dx^*}{dB} = \frac{\partial x^*}{\partial b} \frac{\partial b}{\partial B}
\]
(5.64)

\(^3\)The "if any" part of the question was a dead giveaway. Also, the "at least weakly" tells you this is a Topkis problem straightaway.
So, all we need is for $x^*(b)$ and $y^*(b)$ to be weakly increasing in $b$. So, let's solve for the Marshallian demands.

$$\max_{x,y} \quad x^\alpha y^\beta \quad \text{s.t.} \quad p_x x + p_y y \leq b$$

(5.65)

First-order conditions:

$$\left[ x : \alpha x^{\alpha-1} y^\beta = \frac{\lambda p_x}{\beta x^\alpha y^{\beta-1}} \right] \quad \frac{p_x}{p_y} = \frac{\alpha y}{\beta x} \Rightarrow x p_x = \frac{\alpha}{\beta} y p_y$$

(5.66)

And the constraint will bind, so

$$b = x p_x + y p_y = y p_y \left( 1 + \frac{\alpha}{\beta} \right) = yp_y \left( \frac{\alpha + \beta}{\beta} \right)$$

(5.67)

$$y^*(b) = \frac{\beta b}{p_y (\alpha + \beta)}$$

(5.68)

Symmetrically,

$$x^*(b) = \frac{\alpha b}{p_x (\alpha + \beta)}$$

(5.69)

$\alpha, \beta > 0$, so $x^*$ and $y^*$ are both weakly increasing in $b$.

Hence, no extra assumptions are needed to concluded that $x$ and $y$ are normal goods.

d) First, what does $\frac{\partial^2 v}{\partial x \partial y} > 0$ mean economically? It means that $x$ and $y$ are complements. So, since $p_x$ has increased, we must find that consumption of $x$ decreases in the short run\(^4\). In the long run, consumption of $y$ will decrease, causing $x$ to decrease even more. Hence, tax revenue will decrease from short run to long run.

This is the idea of Le Châtelier: the system reacts to counter change:

Short run

\[ x \downarrow \]

Long run

\[ x \downarrow \downarrow \text{ (even more), } y \downarrow \]

is enough to answer most test questions, but can we prove our intuition mathematically? Consider the short run problem (Walras’ Law is already used to eliminate $z$.)

$$x^*(y, p) = \arg \max_x \left\{ w - p_x x - p_y y + v(x, y) \right\} \equiv \varphi(x, y, p)$$

(5.70)

and the long run problem

$$y^*(p) = \arg \max_y \left\{ w - p_x x^*(y, p) - p_y y + v(x^*(y, p), y) \right\} \equiv \psi(y, p)$$

(5.71)

\(^4\) $h$ must decrease by the law of demand, and since the problem is quasilinear, $x$ will not depend on wealth. With no wealth effects to cancel the substitution effects, the law of demand will hold for $x$ as well as $h$.\n
Now, they have told us about mixed partials, so we suspect that Topkis’ will be involved. Let’s look at the mixed partials of the short run objective, $\varphi$

\[
\frac{\partial^2 \varphi}{\partial x \partial p} = -1 < 0 \quad \text{(5.72)}
\]

\[
\frac{\partial^2 \varphi}{\partial x \partial y} = v_{xy}(x, y) > 0
\]

So, applying Topkis’ theorem, we know

- $x^*(y, p) \searrow$ in $p_x$
- $x^*(y, p) \nearrow$ in $y$

Now, let’s look at the mixed partials of $\psi^5$.

\[
\frac{\partial \psi}{\partial p} = -x^*(y, p) - p_x \frac{\partial x^*}{\partial p}(y, p) + v_x(x^*(y, p), y) \frac{\partial x^*}{\partial p}(y, p) = -x^*(y, p)
\]

\[
\frac{\partial^2 \psi}{\partial p \partial y} = -\frac{\partial x^*}{\partial y}(y, p) > 0 \quad \text{(5.73)}
\]

So, by Topkis’ Theorem, we know

- $y^*(p) \nearrow$ in $p_x$

So, now we can apply our Topkis results.

\[
y^*(p) \geq y^*(p + t)
\]

\[
x^*(y^*(p), p) \geq x^*(y^*(p), p + t) \geq x^*(y^*(p + t), p + t) \quad \text{(5.74)}
\]

or more succinctly

\[
y_{\text{NO TAX}} \geq y_{LR}
\]

\[
x_{\text{NO TAX}} \geq x_{SR} \geq x_{LR} \quad \text{(5.75)}
\]

And since the tax revenue is proportional to $x$, the revenue from the tax will decrease from short run to long run, just as our intuition suggested.

e) What does $\frac{\partial^2 v}{\partial x \partial y} < 0$ mean, economically? It means that $x$ and $y$ are substitutes.

So, here’s our logic:

\[
p_x \nearrow \Rightarrow x_{SR} \searrow
\]

\[
y \text{ is } x \text{'s complement} \Rightarrow y_{LR} \nearrow
\]

\[
y_{LR} \nearrow \text{ makes } x \text{ less effective} \Rightarrow x_{LR} \searrow \searrow
\]

\[
\Rightarrow \text{ tax revenue } \searrow \searrow
\]

More mathematically, if we go through the logic used in the previous part,
we see that the sign of $v_{xy}$ affects how $x^*$ varies with $y$ — now $x^* \downarrow$ in $y$. But this in turn affects how $y^*$ varies with $p_x$ — now $y^* \nearrow$ in $p_x$. These two changes lead to an opposite conclusion about how $y$ changes from short run to long run, but do not affect the conclusion about $x$ decreasing from no tax to short run to long run. Hence, tax revenue still decreases from short run to long run. The system will resist change, or, less mystically, rational maximizer will do whatever they can to avoid being taxed!
Chapter 6

Choice under Uncertainty

6.1 Measures of risk aversion

6.1.1 Coefficient of absolute risk aversion

It is clear that concavity is what leads to risk aversion, as we saw in the class notes (the proof that used Jensen’s inequality). But, why do we use this Arrow-Pratt measure of risk aversion?

Intuition

A risk neutral utility function is linear

\[ u(x) = \alpha + \beta x \] (6.1)

So, let’s start by trying to formulate a measure of risk aversion that is zero-valued for a risk neutral agent. Since \( u''(x) = 0 \) for the aforementioned utility function, we could start with a second derivative. Now, we have to think about normalization of the utility function. Consider the following two utility functions

\[ u(x) = -e^{-x} \quad v(x) = -8e^{-x} \] (6.2)

Both represent the same preferences, but if we only use the second derivative to gauge the level of risk aversion, we would find that \( v(x) \) is much more risk averse. To avoid this problem, we divide by the first derivative.

\[ \frac{u''(x)}{u'(x)} = \frac{v''(x)}{v'(x)} = \frac{-e^{-x}}{e^{-x}} = -1 \] (6.3)

So, the last problem is that the measure of risk aversion we have created thus far has \( U(x) = -e^{-x} \) coming out with a measure that is more negative than that of the risk neutral agent, \( u(x) = \alpha + \beta x \). To make an increase in our measure correspond to an increase in the level of risk aversion, we reorient the scale of the measure by multiplying by \(-1\). This gives us the Arrow-Pratt measure

\[ A(x, u(\cdot)) = -\frac{u''(x)}{u'(x)} \] (6.4)
So, the Arrow-Pratt measure if pretty much the most straightforward way we can measure risk aversion while still avoiding the problems that come with utility being an ordinal concept.

Symmetry of expected utility representations

So, note that the Arrow-Pratt measure gives us a differential equation for the utility; in fact, we can solve this equation in closed form. If we define \( \eta \equiv u' \), then we have

\[
\begin{align*}
A(x) &= -\frac{u''}{u'} = -\frac{\eta'}{\eta} \\
\frac{d\eta}{\eta} &= -A(x)dx \\
\log v &= -\int A(x)dx + \log C \\
u' &= v = C \exp[- \int A(x)dx] \\
u(x) &= C \int^x \exp[- \int^\zeta A(\zeta)d\zeta]d\zeta + D
\end{align*}
\]

where \( C \) and \( D \) are constants of integration. So, for each unique Arrow-Pratt measure, there is a two-dimensional continuum of utility representations that have that measure. This comes from the fact that expected utility representations have two fundamental symmetries – recentering the origin (choosing the \( x \) such that \( u(x) = 0 \)) and linearly rescaling the axes (multiplying \( u(x) \) by a constant).\(^1\)

So, the Arrow-Pratt measure eliminates some of the redundancy due to the invariance symmetries of expected utility representations. Since the mapping between preferences that obey the von Neumann-Morgenstern axioms and Arrow-Pratt measures is one-to-one, it makes sense that we can compare preferences by comparing Arrow-Pratt measures. In contrast, the mapping between second derivatives and preferences that obey the von Neumann-Morgenstern axioms is many-to-one. So, we essentially use Arrow-Pratt to get past what is unimportant and down to something that has one-to-one correspondence with what we care about – the actual preference.

Comparisons across wealth

If \( A(x, u_2(\cdot)) \geq A(x, u_1(\cdot)), \forall x \), we say that \( u_2 \) is more risk averse than \( u_1 \). So, how might we apply this to comparisons across wealth? Let agent 1 have utility function \( u_1(x) \) and let agent 2 have utility function \( u_2(x) \equiv u_1(w + x) \). Agent 1 can be interpreted as an agent with zero wealth, while agent 2 can be interpreted as an agent with wealth level \( w \). If the richer agent, 2, is less risk averse than the poorer agent, 1, then we have \( A(x, u_2(\cdot)) \leq A(x, u_2(\cdot)), \forall x \). Plugging in the definition of \( u_2 \), we arrive at

\[
A(w + x, u_1(\cdot)) \leq A(x, u_1(\cdot)), \forall x
\]

But, this is clearly just equivalent to \( A(x, u_1(\cdot)) \) being decreasing in its first argument. So, we have a simple relation between how \( A(x) \) varies in \( x \) and how the

\(^1\)Note that expected utility representations cannot be rescaled by an arbitrary increasing function like the utility representations we saw earlier in the class. This is because the von Neumann-Morgenstern axioms add extra structure to what utility representations can look like. Concavity of the increasing transform used now matters – namely it must have none.
level of risk aversion changes with wealth. This is why CARA (constant absolute risk aversion – $A(x)$ stays constant) and DARA (decreasing absolute risk aversion) are such useful assumptions. For example, if $A(x)$ is decreasing (if the underlying preferences have DARA), then we know that a wealthier agent will be less risk averse.

### 6.1.2 Coefficient of relative risk aversion

Say that an agent has a wealth level $w$ and is faced with making decisions about proportional gambles, i.e. random variables $t$ where the payoff is $tw$. Then, we can consider the utility function induced by such gambles

$$\tilde{u}(t) = u(tx)$$

(6.7)

So, we measure risk-aversion over choices of $t$ with the Arrow-Pratt measure of induced utility function.

$$A(t, \tilde{u}(\cdot)) = -\frac{\tilde{u}''}{\tilde{u}'} = -\frac{x^2u''(tx)}{xu'(tx)} = -x \frac{u''(tx)}{u'(tx)}$$

(6.8)

We are interested in the Arrow-Pratt measure of $\tilde{u}(t)$ near the actual wealth level, so we set $t = 1$ to get the local coefficient of relative risk aversion

$$A_r(x, u(\cdot)) = -x \frac{u''(x)}{u'(x)}$$

(6.9)

So, we have a nice generalization of the coefficient of absolute risk aversion to risky choices over proportional (relative) bets.

As an aside, which is stronger, DARA of DRRA?

\[
\begin{align*}
\text{DARA} & \implies A'(x) < 0 \\
\text{DRRA} & \implies xA'(x) + A(x) < 0 \\
& \quad A'(x) < -\frac{A(x)}{x}
\end{align*}
\]

(6.10)

Clearly this depends on whether $A(x)$ is positive or negative, i.e. whether the agent in question is a risk-lover or a risk-averter.

### 6.2 Method of proof in insurance/portfolio problems

#### 6.2.1 Insurance

Fully insured?

First off, what does it mean to fully insure? It means that regardless of what happens, an agent will receive the same payoff. In other words, a fully insured agent
has eliminated all risk. Let’s illustrate this thought in the context of the insurance problem from the notes.

Say that there is a probability $p$ of a disaster resulting in a loss $L$. Say that insurance costs $q$ for each dollar of coverage and that it is possible to buy fractions of a policy. Then if $x$ is the level of insurance coverage purchased, then the agent must solve

$$x^* = \arg \max_{x \geq 0} pu(w - L + x - qx) + (1 - p)u(w - qx) \quad (6.11)$$

What level of coverage corresponds to full insurance? Setting the payoffs equal in the two worlds and solving, we find

$$w - L + x - qx = w - qx$$ $$x = L \quad (6.12)$$

So, if $x = L$, then there is no risk in the outcome – the agent will get a payoff of $w - qL$ in both states of the world.

**Out-of-equilibrium reasoning**

Problems concerning uncertainty often use a weird line of reasoning that pertains to the first-order conditions. This logic, which continually comes up in the notes, can best be presented as dealing with an out-of-equilibrium system, i.e. with “first-order conditions” that don’t equal zero.

Consider two optimizations for risk-averse agents, with $u''$, $v''$, $-u'$, $-v' < 0$ and $c', c'' > 0$.

$$x^* = \arg \max u(x) - c(x) \quad \text{and} \quad \tilde{x} = \arg \max v(x) - c(x)$$

$$[x^*]: \frac{\partial u}{\partial x}(x^*) - \frac{\partial c}{\partial x}(x^*) = 0 \quad \quad [\tilde{x}]: \frac{\partial v}{\partial x}(\tilde{x}) - \frac{\partial c}{\partial x}(\tilde{x}) = 0 \quad (6.13)$$

So, at the appropriate maximizer, the first derivative of the objective function is equal to zero. But, what if we plug the wrong maximizer in? Say we plug $x^*$ into the $\sim$ objective, and that we somehow conclude that

$$\frac{\partial v}{\partial x}(x^*) - \frac{\partial c}{\partial x}(x^*) > 0 \quad (6.14)$$

What does this tell us about the relationship between $x^*$ and $\tilde{x}$? It depends on how the left-hand side of the inequality changes in $x$. Define the left-hand side to be $\varphi(x^*)$. Then, the derivative of $\varphi(x)$ is given by

$$\frac{\partial \varphi}{\partial x}(x) = \frac{\partial^2 v}{\partial x^2}(x) - \frac{\partial^2 c}{\partial x^2}(x) < 0, \forall x \quad (6.15)$$

where the last inequality comes from the original assumptions we made about $v$ and $c$. Hence $\varphi(x)$ is a downward sloping function. From first-order conditions, we have that $\varphi(\tilde{x}) = 0$. We have already found (somehow) that $\varphi(x^*) > 0$. Hence, we must conclude that $x^* < \tilde{x}$. This is the big idea of most problems that relate to choice under uncertainty. If a first-order condition gets knocked away from zero, then what must we do to return it to its equilibrium (zero) value?
Now, we can turn back to the insurance problem. Our first-order condition is

$$p(1-q)u'(w - L + (1-q)x^*) - q(1-p)u'(w - qx^*) \equiv \frac{\partial U}{\partial x}(x^*) = 0 \quad (6.16)$$

Now, if insurance is \textbf{actuarially fair}, i.e. if $p = q$, then this condition becomes

$$u'(w - L + (1-p)x^*) = u'(w - px^*) \quad (6.17)$$

And since $u'$ is monotonic$^2$, we have

$$w - L + (1-p)x^* = w - px^*$$

$$x^* = L \quad (6.18)$$

So, the agent fully insures. But, we could have done this even without first-order conditions. Say we choose some insurance level $x$. Then $p = q$ tells us that the expected payoff is

$$\mathbb{E} [\text{payoff}] = p(w - L + (1-p)x) + (1-p)(w - px) = pw - pL + p(1-p)x + (1-p)w - p(1-p)x$$

$$= w - pL \quad (6.19)$$

So, the expected payoff is the same, regardless of the insurance coverage level chosen. Since $x = L$ makes the payoff in both states the same, all other insurance levels give distributions across states that are a mean-preserving spread of the $x = L$ level. Hence, a risk-averse agent will prefer to set $x = L$. Note that we did not say that the agent simply minimizes the variance of his outcome. If two lotteries have the same mean, it is \textit{not} always true that the agent will take the lottery with the lowest variance. Agents can care about all moments of the distribution, not just about the mean and variance of it. So be careful to only make conclusions based on mean-preserving spreads and not merely on relative variances.

Now, what if $p > q$? Then, the first-order conditions (6.16) can be arranged to yield

$$1 < \frac{p}{q} \frac{1-q}{1-p} = \frac{u'(w - qx^*)}{u'(w - L + (1-q)x^*)} \quad (6.20)$$

$^2$This assures us of a well-defined inverse, $u^{-1}$. 
where the inequality comes from the fact that if \( p/q > 1 \), then \( \frac{1}{1-p} > 1 \) as well. And, since \( u' \) is decreasing, we must have

\[
\frac{w - qx^*}{L} < \frac{w - L + x^* - qx^*}{x^*}
\]

(6.21)

So, the agent will overinsure. He sees that insurance is cheap and actually hopes for the disaster to happen! Through a similar chain of logic, we see that if \( p < q \), the insurance is too expensive and the agent, hoping that disaster is averted, underinsures.

**Wealth effects**

So, how will the optimal level of insurance coverage change with wealth? Let’s look at the wealth derivative of \( \frac{\partial U}{\partial x} \).

\[
\frac{\partial^2 U}{\partial x \partial w} = p(1-q)u''(w - L + (1-q)x) - q(1-p)u''(w - qx)
\]

(6.22)

We know that wealth effects are the product of CARA/DARA/IARA type assumptions. So, how can we get the Arrow-Pratt measure into the mix? Note that the first-order condition \( \frac{\partial U}{\partial x}(x^*) = 0 \) tells us that

\[
p(1-q)u'(w - L + (1-q)x^*) = q(1-p)u'(w - qx^*)
\]

(6.23)

So, we rewrite \( \frac{\partial^2 U}{\partial x \partial w}(x^*) \) with the freedom introduced by the first-order condition by dividing the firm term of (6.22) by the left hand side of (6.23) and the second term of (6.22) by the right hand side of (6.23), yielding

\[
\frac{\partial^2 U}{\partial x \partial w}(x^*) = \frac{p(1-q)u'(w - L + (1-q)x^*)}{u'(w - L + (1-q)x^*)} \left[ u''(w - L + (1-q)x) \frac{u''(w - L + (1-q)x^*)}{u'((w - q)x)} \right]
\]

(6.24)

So, plugging in the definition of the coefficient of absolute risk aversion yields

\[
\text{sgn} \left[ \frac{\partial^2 U}{\partial x \partial w}(x^*) \right] = \text{sgn}[A(w - qx^*) - A(w - qx^* + (x^* - L))]
\]

(6.25)

So, given the sign of \( x^* - L \), we can determine the sign of \( \frac{\partial^2 U}{\partial x \partial w}(x^*) \) if we have a DARA type assumption.

For instance, if \( p < q \), the we underinsure, i.e. \( x^* < L \). If we have DARA, then we can sign \( \frac{\partial^2 U}{\partial x \partial w}(x^*) \leq 0 \). The chain of logic is as follows:

- At \( x^*(w) \), \( \frac{\partial U}{\partial x}(x^*) = 0 \)
- But, \( \frac{\partial U}{\partial w}(\frac{\partial U}{\partial x}(x^*)) \leq 0 \), so \( \frac{\partial U}{\partial x} \) is decreasing in wealth at \( x^*(w) \)
- So, if we increase \( w \) while holding \( x^* \) constant, then we will get \( \frac{\partial U}{\partial x} \leq 0 \). How must we change \( x^* \) to re-equilibrate (i.e. attain \( \frac{\partial U}{\partial x} = 0 \))?
• \( \frac{\partial}{\partial x} \left[ \frac{\partial U}{\partial x} \right](x^*) = \frac{\partial^2 U}{\partial x^2}(x^*)p(1-q)^2u''(w - L + (1 - q)x) + q^2(1-p)u''(w - qx) < 0 \) (since \( u \) is concave)

• So, we decrease \( x^* \) to pull \( \frac{\partial U}{\partial x} \) back down to zero.

• Hence, \( x^*(w) \) is decreasing if we have DARA.

The logic is basically of the form, “Say I treat \( w \) and \( x^* \) independently. Increasing \( w \) decreases \( \frac{\partial U}{\partial x} \) while increasing \( x^* \) increases \( \frac{\partial U}{\partial x} \). So, if \( x^* \) is functionally related to \( w \) in a way such that \( \frac{\partial U}{\partial x}(x^*, w) = \frac{\partial U}{\partial x}(x^*(w)) = 0 \), then \( x^*(w) \) will have to move in opposition to \( w \), i.e. \( x^*(w) \) will be decreasing.”

For a “more math, less intuition” approach to this problem, note that the first-order condition requires

\[
\frac{\partial U}{\partial x}(x^*(w), w) = 0 \tag{6.26}
\]

It defines \( x^*(w) \) implicitly. Taking a wealth derivative of both sides yields

\[
\frac{\partial}{\partial w} \left[ \frac{\partial U}{\partial x}(x^*(w), w) \right] = \frac{\partial^2 U}{\partial x^2}(x^*(w), w) \cdot \frac{\partial x^*}{\partial w}(w) + \frac{\partial^2 U}{\partial x \partial w}(x^*(w), w) = 0 \tag{6.27}
\]

and hence

\[
\frac{\partial x^*}{\partial w}(w) = -\frac{\frac{\partial^2 U}{\partial x^2}(x^*(w), w)}{\frac{\partial^2 U}{\partial x \partial w}(x^*(w), w)} \tag{6.28}
\]

At optimum, the second-order conditions tell us that

\[
\frac{\partial^2 U}{\partial x^2}(x^*(w), w) \leq 0 \tag{6.29}
\]

So, we find that, in the end,

\[
\operatorname{sgn} \left[ \frac{\partial x^*}{\partial w}(w) \right] = \operatorname{sgn} \left[ \frac{\partial^2 U}{\partial x \partial w}(x^*(w), w) \right] \tag{6.30}
\]

which leads to the same conclusion as the more intuitive logic presented above.

6.2.2 Portfolio

Is no investment possible?

Consider an agent who must choose a fraction of her wealth to invest in a risky asset with stochastic return \( z \) and a fraction to invest in a risk-free asset with return \( r \). She must solve

\[
\max_{0 \leq a \leq w} \int_{\frac{z}{\beta}}^{\frac{1}{\beta} + \frac{1}{\mu} \lambda} u(az + (w - a)r) dF(z) \tag{6.31}
\]

As a benchmark, let’s first consider a risk-neutral agent with Bernoulli utility \( u(x) = \alpha + \beta x \). First-order conditions then give us

\[
\mathbb{E}[z] - r = \frac{1}{\beta} (\lambda - \mu) \tag{6.32}
\]
Now, it is impossible to have an \( a \) that hits both boundaries of the constraint set at once, so either \( \lambda \) or \( \mu \) must be zero-valued. If \( \mathbb{E} [z] > r \), then we must have \( \mu = 0 \) and \( \lambda \geq 0 \). Hence, the \( \lambda \) constraint must bind and we will get \( a^* = w \). Similarly, if \( \mathbb{E} [z] < r \), then we must have \( \mu > 0 \) and \( \lambda = 0 \), yielding the conclusion that the \( \lambda \) constraint binds, i.e. that \( a^* = 0 \).

So, a risk-neutral agent simply invests all of his wealth in the asset with the best return. What happens though, if \( u \) is a nice concave (risk-averse) function as well? The, the first order conditions tell us that

\[
\frac{\partial U}{\partial a}(a^*) = \int u'(a(z - r) + wr)(z - r)dF(z) = \lambda - \mu
\]  

(6.33)

If \( a^* = 0 \), then we must have that \( \lambda = 0 \), which tells us that

\[
\frac{\partial U}{\partial a}(0) = \int u'(wr)(z - r)dF(z) = u'(wr)(\mathbb{E} [z] - r) \leq 0
\]  

(6.34)

Since \( u' > 0 \), it must be the case that \( \mathbb{E} [z] \leq r \). Hence, if \( \mathbb{E} [z] > r \), we must conclude that \( a^* = 0 \) is suboptimal. So, for any asset with an expected return greater than the risk-free rate of return, the agent must invest a strictly positive amount in the risk asset.

**Comparative statics of risk**

Say we have two agents, \( u \) and \( v \). Let \( u \) be the more risk-averse\(^3\) of the two. Consider

\[
a^* = \arg \max_a \int u[a(z - r) + wr]dF(z)
\]

\[
\tilde{a} = \arg \max_a \int v[a(z - r) + wr]dF(z)
\]  

(6.35)

How does \( a^* \) compare to \( \tilde{a} \)? The first-order conditions for the * problem state that

\[
\int u'(a^*(z - r) + wr)(z - r)dF(z) = 0
\]  

(6.36)

If \( u \) is more risk averse than \( v \), then we know that \( v = h \circ u \), where \( h \) is a non-decreasing and convex function (a de-concavifier, if you will). To see this, define \( y \equiv u(x) \). Then, \( x = u^{-1}(y) \). We know that this inverse exists because \( u \) is strictly increasing. So, by our definition of \( h \), \( h(y) = v(u^{-1}(y)) \). \( h \) is well-defined, so now we just need to show that it is increasing and convex. Taking a derivative,

\[
h'(y) = v'(u^{-1}(y))\frac{\partial u^{-1}}{\partial y}(y) = \frac{v'(u^{-1}(y))}{u'(u^{-1}(y))} > 0
\]  

(6.37)

where the last inequality comes from differentiating the identity \( u(u^{-1}(y)) = y \), and the inequality comes from the assumptions we made about \( u \) and \( v \). So, \( h \) is increasing. What about the concavity of \( h \)? Differentiating again, we find

\[
h''(y) = \frac{v''(u^{-1}(y))}{u'(u^{-1}(y))}\frac{\partial u^{-1}}{\partial y}(y) - \frac{v'(u^{-1}(y))}{[u'(u^{-1}(y))]^2}\frac{u''(u^{-1}(y))}{\partial y}(y)
\]  

(6.38)

\(^3\)In other words, \( -\frac{v''(x)}{v'(x)} \leq \frac{u''(x)}{u'(x)}, \forall x \).
or
\[
\frac{h''(y)}{[u'(u^{-1}(y))]^2} \left\{ v''(u^{-1}(y)) - \frac{v'(u^{-1}(y))}{u'(u^{-1}(y))} u''(u^{-1}(y)) \right\} \tag{6.39}
\]
So, the condition for \( h \) to be convex \((h'' > 0)\) is that the expression in the curly brackets must be positive. But, rearranged, this just says
\[
-\frac{v''(u^{-1}(y))}{v'(u^{-1}(y))} \leq -\frac{u''(u^{-1}(y))}{u'(u^{-1}(y))} \tag{6.40}
\]
which is the definition of \( u \) being more risk averse than \( v \). So, if \( u \) is more risk averse than \( v \), then we can construct a convex, increasing \( h \) such that \( v(x) = (h \circ u)(x) \).

So, call the objective in the \( \sim \) problem \( \bar{\varphi} \) and the objective in the \( * \) problem \( \varphi^* \). Then, using the \( h \) we just constructed, we find
\[
\frac{\partial \bar{\varphi}}{\partial a}(\tilde{a}) = \int v'[a(z-r) + wr](z-r)dF(z)
\]
\[
= \int h'[u(a(z-r) + wr)] \cdot u'[a(z-r) + wr](z-r)dF(z)
\]
and
\[
\frac{\partial \varphi^*}{\partial a}(a^*) = \int u'[a(z-r) + wr](z-r)dF(z)
\]
Notice that these two derivatives are quite similar. \( \frac{\partial \bar{\varphi}}{\partial a} \) is just the same as \( \frac{\partial \varphi^*}{\partial a} \) except that the integrand is multiplied by a funny \( h' \) term. How will this change the integral? Well, since \( h' \) is increasing in \( z \) for all \( a \) and for all parameter values, it will place more “probability weight” on \( z \) values that are higher.

Now, note that the first order condition for the \( \sim \) problem states that
\[
\frac{\partial \bar{\varphi}}{\partial a}(\tilde{a}) = 0 \tag{6.42}
\]
and the first order condition for the \( * \) problem states that
\[
\frac{\partial \varphi^*}{\partial a}(a^*) = \int u'[a^*(z-r) + wr](z-r)dF(z) = 0 \tag{6.43}
\]
Note that this integrand is positive for all \( z > r \) and negative for all \( z < r \) (since \( u' \) is positive everywhere). Geometrically, then, we can interpret this first order condition to mean that if we plot \( u'[a^*(z-r) + wr](z-r)f(z) \equiv \psi \), then the area between the curve and the \( z \)-axis for \( z < r \) is exactly equal to the area between the curve and the \( z \)-axis for \( z > r \).
So, what if we plug the * optimizer into the derivative of the \( \sim \) objective instead. It will clearly throw off the delicate balance of areas illustrated above. Since, as discussed above, the extra \( h' \) multiplier places more relative weight on higher \( z \) values in the integral, we must conclude that if we were to plot \( u[a^*(z - r) + wr] \cdot h'[u(a^*(z - r) + wr)]f(z) \), the corresponding “B” area would be larger than the corresponding “A” area. Since the value of the integral is just \( B - A \), we then conclude that

\[
\frac{\partial \tilde{\varphi}}{\partial a}(a^*) > 0 \quad (6.44)
\]

Now, to relate \( a^* \) to \( \tilde{a} \), we just have to know how \( \frac{\partial \tilde{\varphi}}{\partial a} \) varies in \( a \). Taking a derivative, we have

\[
\frac{\partial \tilde{\varphi}}{\partial a}(a) = \int u''[a(z - r) + wr](z - r)^2 h'[u(a^*(z - r) + wr)]dF(z) < 0 \quad (6.45)
\]

since \( u'' \) is always negative, \( h' \) is always positive, and anything squared is always positive. So, if \( \frac{\partial \tilde{\varphi}}{\partial a} \) is decreasing in \( a \), and \( \frac{\partial \tilde{\varphi}}{\partial a}(a^*) > \frac{\partial \tilde{\varphi}}{\partial a}(\tilde{a}) = 0 \), then we must conclude that \( a^* < \tilde{a} \).

**Comparative statics of wealth**

With wealth effects, we essentially have the same mess.

\[
\frac{\partial^2 U}{\partial a \partial w}(a) = \int ru''[a(z - r) + wr](z - r)dF(z) \quad (6.46)
\]

How can we get Arrow-Pratt into the mix? Divide and multiply by \( u' \).

\[
\frac{\partial^2 U}{\partial a \partial w}(a) = \int ru''[a(z - r) + wr]u'[a(z - r) + wr](z - r)dF(z) \quad (6.47)
\]

Say that we have DARA. It tells us that \( A(x) \) is decreasing. So, using \(-A(x)\) as a weighting function, we see that again, we are placing relatively more weight on higher values of \( z \), and that the integrand is positive for \( z > r \) and negative for \( z < r \). Hence, we find that

\[
\frac{\partial^2 U}{\partial a \partial w}(a) > 0 \quad (6.48)
\]

So, \( a^*(w) \) is increasing if we have DARA.

### 6.3 Portfolio theory applied

#### 6.3.1 Guiding question (#3 from the 2004 202 Final)

Ann is an expected utility maximizing individual with a strictly increasing, strictly concave, and twice differentiable Bernoulli utility function \( u(\cdot) \) defined over wealth. She has initial wealth \( w \). One day she goes to the racetrack and has the opportunity to bet on a race with \( n \geq 2 \) horses, \( i = 1, \ldots, n \). Ann assesses the probabilities of
horse \( i \) winning as \( p_i \), with \( \sum_i p_i = 1 \). A ticket that pays 1 in the event that horse \( i \) wins costs \( r_i \), with \( \sum_i r_i = 1 \). Ann can buy as many tickets as she likes (assume that she can buy ticket fractions and also that she never wants to spend more than her whole initial wealth) taking the prices as fixed. Let \( x_i \) denote her purchases.

(a) Write Ann’s optimization problem.

(b) Show that if \( p_i/r_i > p_j/r_j \), then Ann will bet more on horse \( i \) than on horse \( j \).

(c) Show that if \( p_i/r_i > 1 \), then Ann will gamble a strictly positive amount on horse \( i \).

Assume that \( n = 2 \) and that Ann has DARA preferences for the rest of the problem.

(d) Show that her payoffs in the two possible states of the world only depend on \( \Delta \equiv x_1 - x_2 \).

(e) Show how \( \Delta \) will change with her initial wealth.

6.3.2 Solution

(a) Define \( w_i \equiv w - \sum_k r_k x_k + x_i \) and \( U(x) \equiv \sum_i p_i u(w_i) \). Then, Ann must solve

\[
\max_{x \geq 0} U(x) \quad (6.49)
\]

(b) So, let’s consider how \( U(x) \) moves in \( x_i \) and \( x_j \). Note that for an interior solution, these derivatives must be equal to zero, and for a corner solution (\( x_i = 0 \)), the derivative must be less than zero

\[
\frac{\partial U}{\partial x_i}(x^*) = \sum_k p_k r_k u'(w_k) + p_i u'(w_i) \leq 0
\]

\[
\frac{\partial U}{\partial x_j}(x^*) = \sum_k p_k r_j u'(w_k) + p_j u'(w_j) \leq 0 \quad (6.50)
\]

Now, say that \( x_j > x_i \geq 0 \). Then, we have \( \frac{\partial U}{\partial x_j}(x^*) = 0 \). Hence,

\[
\sum_k p_k r_j u'(w_k) = p_j u'(w_j) \quad (6.51)
\]

Plugging this into the inequality for \( \frac{\partial U}{\partial x_i}(x^*) \), we find

\[
\frac{p_i}{r_i} u'(w_i) \leq \frac{p_j}{r_j} u'(w_j) \quad (6.52)
\]

Since \( x_j > x_i \), we can say that \( w_j > w_i \). Then, since \( u' \) is decreasing, we find that \( u'(w_j) < u'(w_i) \). If this is true, as well as \( p_i/r_i > p_j/r_j \), then (6.52) cannot hold.

So, we have that if \( p_i/r_i > p_j/r_j \), then Ann must bet more on horse \( i \) than on horse \( j \).
(c) Say that there is a horse $i$ with $p_i/r_i > 1$ and $x_i = 0$. First, note that Ann does worst in states where a horse wins that she did not bet on, and that her wealth is the same in all such states. This worst wealth, call it $\overline{w}$ is given by

$$\overline{w} = w - \sum_k r_k x_k$$  \hspace{1cm} (6.53)

So, $w_j \geq \overline{w}$ with equality holding only in $x_j = 0$. Now, we consider how $U(x)$ changes with the bet on our horse $i$. Since $x_i = 0$, $w_i = \overline{w}$, we have

$$\frac{\partial U}{\partial x_i}(x) = p_i u'(w) - r_i \sum_k p_k u'(w_k)$$

$$\frac{1}{r_i u'(w)} \frac{\partial U}{\partial x_i}(x) = \left\{ \begin{array}{ll}
\frac{p_i}{r_i} & 1 < 1, \text{since } \sum p_k = 1 \\
> 1, & 1 < 1
\end{array} \right.$$  \hspace{1cm} (6.54)

which contradicts optimality. So, if $p_i/r_i > 1$, we have that Ann prefers to bet a non-zero amount on horse $i$.

(d) First, note that in the case $n = 2$, we have

$$r_1 = 1 - r_2 \equiv r$$  \hspace{1cm} (6.55)

$$p_1 = 1 - p_2 \equiv p$$  \hspace{1cm} (6.56)

The wealth levels in the two states of the world are then given by

$$w_1 = w + x_1 - rx_1 - (1-r)x_2 = (1-r)(x_1 - x_2) = (1-r)\Delta$$

$$w_2 = w + x_2 - rx_1 - (1-r)x_2 = r(x_2 - x_1) = -r\Delta$$  \hspace{1cm} (6.57)

So, the outcome of the betting strategy only depends on $\Delta$.

(e) The objective for this problem is

$$U(\Delta) = pu[w + (1-r)\Delta] + (1-p)u[w - r\Delta]$$  \hspace{1cm} (6.58)

So, the first order condition is

$$p(1-r)u'[w + (1-r)\Delta] = r(1-p)u'[w - r\Delta]$$  \hspace{1cm} (6.59)

So, what if $p = r$? Then, the expected return on both horses is the same. Using this, we find

$$u'[w + (1-r)\Delta] = u'[w - r\Delta]$$  \hspace{1cm} (6.60)

And since $u'$ is decreasing, we have, in the case $p = r$, that $\Delta = 0$, independent of wealth.

Now we consider the more interesting case. Say $\frac{p}{r} > \frac{1-p}{1-r}$. Then, necessarily $\frac{p}{r} > 1$. By (c), we then know that $x_1$ is strictly positive, and by (b) we know
that $x_1 > x_2$. So, $\Delta > 0$. Dividing the first order condition by $r(1 - r)$, we find

$$\frac{p}{r} u'[w + (1 - r)\Delta] = \frac{1 - p}{1 - r} u'[w - r\Delta]$$

(6.61)

Rearranging, we then attain

$$\frac{p}{r} \frac{1 - r}{1 - p} = \frac{u'[w - r\Delta]}{u'[w + (1 - r)\Delta]} \equiv \varphi$$

(6.62)

So, the left hand side certainly does not change with wealth. So, we need

$$\frac{\partial \varphi}{\partial w} = 0.$$ This implies

$$\frac{\partial \varphi}{\partial w} = \frac{u'[w - r\Delta]}{u[w + (1 - r)\Delta]}$$

(6.63)

Rearranging,

$$A(w - r\Delta) \left( 1 - r \frac{\partial \Delta}{\partial w} \right) = A(w + (1 - r)\Delta) \left( 1 + (1 - r) \frac{\partial \Delta}{\partial w} \right)$$

(6.64)

or

$$A(w - r\Delta) - A(w + (1 - r)\Delta) = \frac{\partial \Delta}{\partial w} \left\{ (1 - r)A(w + (1 - r)\Delta) + rA(w - r\Delta) \right\}$$

> 0, by DARA

(6.65)

Thus, $\frac{\partial \Delta}{\partial w} > 0$.

### 6.4 Precautionary savings

#### 6.4.1 Guiding question (MWG 6.C.9)

The purpose of this problem is to examine the implications of uncertainty and precaution in a simple consumption-savings problem.

In a two-period economy, a consumer has first-period initial wealth $w$. The consumer’s utility level is given by

$$U(c_1, c_2) = u(c_1) + v(c_2)$$

(6.66)

where $u(\cdot)$ and $v(\cdot)$ are concave functions, and $c_1$ and $c_2$ denote consumption levels in the first and second period, respectively. Denote by $x$ the amount saved by the consumer in the first period (so that $c_1 = w - x$ and $c_2 = x$), and let $x_0$ be the optimal value of $x$ in this problem.

We now introduce uncertainty into this economy. If the consumer saves as amount $x$ in the first period, his wealth in the second period is given by $x + y$, where $y$ is distributed according to $F(\cdot)$. In what follows, $E[\cdot]$ always denotes the expectation with respect to $F(\cdot)$. Assume that the Bernoulli utility function over realized wealth levels in the two periods ($w_1, w_2$) is $u(w_1) + v(w_2)$. Hence, the consumer now solves

$$\max_x u(w - x) + E[v(x + y)]$$

(6.67)

Denote the solution to this problem by $x^*$. 
(a) Show that if $E[v'(x_0 + y)] > v'(x_0)$, then $x^* > x_0$.

(b) Define the coefficient of absolute prudence of a utility function to be $-v''''(x)/v''(x)$. Show that if the coefficient of absolute prudence of a utility function $v_1(\cdot)$ is not larger than the coefficient of absolute prudence of utility function $v_2(\cdot)$ for all levels of wealth, then $E[v'_1(x_0 + y)] > v'_1(x_0)$ implies $E[v'_2(x_0 + y)] > v'_2(x_0)$. What are the implications of the fact in the context of part a)?

(c) Show that if $v'''(\cdot) > 0$ and $E[y] = 0$, then $E[v'(x + y)] > v'(x)$ for all values of $x$.

(d) Show that if the coefficient of absolute risk aversion of $v(\cdot)$ is decreasing with wealth, then $-v''''(x)/v''(x) > -v''(x)/v'(x)$ for all $x$, and hence $v'''(\cdot) > 0$.

6.4.2 Solution

(a) Consider the two problems we are to compare

$$x_0 = \arg\max_x u(w - x) + v(x)$$

$$x^* = \arg\max_x u(w - x) + E[v'(x + y)]$$

(6.68)

The first-order condition for the $x_0$ problem is

$$u'(w - x_0) = v'(x_0)$$

(6.69)

How is $\varphi^*$ acting at $x_0$?

$$\frac{\partial \varphi^*}{\partial x}(x_0) = -u'(w - x_0) + E[v'(x_0 + y)]$$

(6.70)

At $x_0$, we can use the first-order conditions to get

$$\frac{\partial \varphi^*}{\partial x}(x_0) = -v'(x_0) + E[v'(x_0 + y)]$$

(6.71)

So, if we have $E[v'(x_0 + y)] > v'(x_0)$, then we also have $\frac{\partial \varphi^*}{\partial x}(x_0) > 0 = pd\varphi^* x(x^*)$. So, to rank $x_0$ and $x^*$, we will need to put a sign on the derivative of $\frac{\partial \varphi^*}{\partial x}(x)$. Calculating

$$\frac{\partial^2 \varphi^*}{\partial x^2}(x) = u''(w - x) + E[v''(x_0 + y)] < 0$$

(6.72)

where the inequality holds because of the concavity of the functions and the fact that the expectation of a random variable that is always negative must itself be negative.

So, (6.71) then tells us that $x^* > x_0$. 
(b) This prudence measure looks a lot like the Arrow-Pratt measure. Let’s capitalize on this
\[ \eta_1(x) = -v_1'(x) \]
\[ \eta_2(x) = -v_2'(x) \] (6.73)

Note that the negative signs make our \( \eta \)'s increasing. Arrow-Pratt is then given by
\[ A(x, \eta_1(\cdot)) = -\frac{\eta''_1}{\eta'_1} = -\frac{v'''_1}{v''_1} \] (6.74)

So, \( \eta \)'s coefficient of absolute risk aversion is equivalent to \( v \)'s coefficient of absolute prudence. Hence
\[ -\frac{v'''_1}{v''_1} \leq -\frac{v'''_2}{v''_2} \iff -\frac{\eta''_1}{\eta'_1} \leq -\frac{\eta''_2}{\eta'_2} \] (6.75)

Now, we are trying to prove that if \( -\frac{v'''_1}{v''_1} \leq -\frac{v'''_2}{v''_2} \) holds, then
\[ \mathbb{E}[v_1'(x_0 + y)] > v_1'(x_0) \Rightarrow \mathbb{E}[v_2'(x_0 + y)] > v_2'(x_0) \] (6.76)

Mapping this statement into a statement in terms of the \( \eta \)'s, we have that if \( -\frac{\eta''_1}{\eta'_1} \leq -\frac{\eta''_2}{\eta'_2} \) then
\[ \mathbb{E}[\eta'_1(x_0 + y)] < \eta'_1(x_0) \Rightarrow \mathbb{E}[\eta'_2(x_0 + y)] < \eta'_2(x_0) \] (6.77)

But, this is clearly true. If \( \eta_1 \) is less risk-averse than \( \eta_2 \), the if \( \eta_1 \) refuses a bet, so will \( \eta_2 \), by definition. So, the same statement in terms of the \( v \)'s will hold as well.

So, what does this tell us about part (a)? If 1 is induced to save more by the risk in \( y \), then so is 2 if
\[ -\frac{v'''_1}{v''_1} \leq -\frac{v'''_2}{v''_2} \] (6.78)

What's more, 2 will save even more than 1. Say that \( v_1 \) is more prudent than \( v_2 \) in the following problems
\[ x^* = \arg \max_x u(w - x) + \mathbb{E}[v_1(x + y)] \]
\[ \tilde{x} = \arg \max_x u(w - x) + \mathbb{E}[v_2(x + y)] \] (6.79)

This is the same as \( v_1' \) being more risk-averse than \( v_2' \). This means that the certain equivalent\(^4\) of \( v_1' \) for a certain lottery is higher than that of \( v_2' \). But, since \( v'''_1, v'''_2 < 0 \), this means that the expected value of a lottery for \( v_1 \) is less than it is for \( v_2 \). Hence, we have
\[ \mathbb{E}[v_1'(x + y)] \leq \mathbb{E}[v_2'(x + y)] \] (6.80)

\(^4\)The certain equivalent of the lottery \( x + y \) for \( v_1' \) is defined implicitly by \( v_1'(c_1) = \mathbb{E}[v_1'(x + y)] \).
Now, we know that $\frac{\partial \varphi^*}{\partial x}(x^*) = 0$ by the first-order conditions of the * problem. If we plug $\tilde{x}$ into this derivative, then we can write

$$\frac{\partial \varphi^*}{\partial x}(\tilde{x}) = \frac{\partial \tilde{\varphi}}{\partial x}(\tilde{x}) + \mathbb{E} [v_1'(x + y)] - \mathbb{E} [v_2'(x + y)] < 0 \quad (6.81)$$

So,

$$\frac{\partial \varphi^*}{\partial x}(\tilde{x}) < \frac{\partial \varphi^*}{\partial x}(x^*) = 0 \quad (6.82)$$

and we have already shown that $\frac{\partial^2 \varphi^*}{\partial x^2}(x) < 0$. Hence, we find that $x^* < \tilde{x}$, i.e. that the more prudent agent will save more in anticipation of the uncertainty in the future that is encapsulated in $y$.

So, this prudence coefficient tells us about precautionary savings. It seems like this should be a function of the Arrow-Pratt measure, but it is not. $u'''$ has an important role to play in risk analysis.

(c) If $u''' > 0$, then $\eta'' = -u''' < 0$, so $\eta(x)$ is concave. Now, since $\mathbb{E} [y] = 0$, we know that $x + y$ is a mean-preserving spread of $x$. Hence, by Jensen’s inequality (or less technically, risk aversion), we have

$$\mathbb{E} [\eta(x + y)] \leq \eta(x) \quad (6.83)$$

or

$$\mathbb{E} [v'(x + y)] \geq v'(x) \quad (6.84)$$

(d)

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

$$A'(x) = -\frac{v'''}{v''} + \left(\frac{v'''}{v''}\right)^2 = \frac{v''}{v'} \left( -\frac{v'''}{v''} + \frac{v''}{v'} \right) \quad (6.85)$$

where the last inequality is what must hold if $A(x)$ is decreasing. So, since $A(x) > 0$ for a risk-avertor, we will need $\frac{v'''}{v''} < \frac{v''}{v'}$ for the inequality to hold, or negating,

$$-\frac{v'''}{v''} > -\frac{v''}{v'} \quad (6.86)$$

Inserting the sign of each element of this inequality, we have

$$-\frac{v'''}{(-)} > -\frac{v''}{(+)} \quad (6.87)$$

If $v''' < 0$, then this can never hold. Hence, $v''' > 0$.

So, we have shown that DARA $\Rightarrow v''' > 0$. 
Chapter 7

General Equilibrium I

7.1 Hemicontinuity and fixed-point theorems

7.1.1 Correspondences

In class, we proved the existence of Walrasian equilibrium through the use of somewhat complicated ideas concerning continuity of correspondences. So, let’s talk about these ideas, at least superficially. To start, we must ask, “What is a correspondence, exactly?”

A correspondence is a generalization of a function that maps to sets instead of individual points. In terms of what we are talking about today these correspondences will be of the form

\[ f : A \mapsto \mathcal{P}(A) \]  

(7.1)

where \( \mathcal{P}(A) \) is the power set of \( A \), or the set of all subsets of \( A \) (sometimes denoted \( 2^A \)). We often write this more simply as

\[ f : A \Rightarrow A \]  

(7.2)

So, when would we use correspondences? Most commonly, they come up when we have multiple solutions to an optimization. For instance, consider the Marshallian demand problem:

\[
x^*(p, w) = \arg \max_{x \geq 0} u(x) \text{ s.t. } p \cdot x \leq w
\]  

(7.3)

If there are multiple \( x \)'s that lead to the same optimum value, then the Marshallian demand is no longer a function; it is a correspondence.

7.1.2 Hemicontinuity

So, how might we generalize the idea of continuity to correspondences? Standard \( \varepsilon - \delta \) notions of continuity seem like they would have a hard time generalizing, since \( f(x) \) could be a set that cannot fit inside of an arbitrarily small \( \varepsilon \)-neighborhood. Perhaps we would be better off with a notion of continuity that depends on sequences and limit points.
Figure 7.1: Are these graphs closed?

But first, let’s start with a little notation. The graph of a correspondence $f$ is the set of points $\{(x, y) | y \in f(x)\}$. A correspondence $f$ has a closed graph if given a convergent sequence in $A$, $\{x_n\} \to x$, such that $x \in A$ and a convergent sequence $\{y_n\} \to y$, such that $y_n \in f(x_n), \forall n$, we have $y \in f(x)$.

This is the standard notion of a closed set applied to the graph of the correspondence – in other words, it is exactly the definition we expect from the name. Let’s test out the idea. Figure 7.1(a) is clearly closed, but what about (b)? The arrow on the graph points out a likely objection. What about this sequence? Here, the $y$ sequence is unbounded – $\lim_{m \to \infty} y_m$ does not exist. Hence (b) does not violate our definition. It simply consists of two closed (but not connected) sets. (Remember, $[a, \infty)$ is closed!)

So, how might we think of extending the idea of continuity to correspondences? Using the idea of a sequence-based definition, we start with a sequence of values in the domain, $\{x_n\} \to x$. Given these $x$-coordinates of our path, we want a nice set of $y$-coordinates that make a well-behaved path inside the graph. These are the notions that will be formalized in our discussion of upper and lower hemicontinuity.

A correspondence is upper hemicontinuous if it has a closed graph and compact sets have bounded images. So (a) is uhc, while (b) is not (a neighborhood around the dashed line does not have a bounded image). In words, upper hemicontinuity states that, given a convergent path in the graph of a correspondence, the limit point of that path must also be in the correspondence.

Still, (a) doesn’t seem like it should be considered continuous per se. Let’s introduce another notion of correspondence continuity – that of lower hemicontinuity (lhc).

A correspondence is lower hemicontinuous if compact sets have bounded images, and if, when we have a convergent sequences in the domain $\{x^m\} \to x \in A$, we can find, for all $y \in f(x)$ a convergent sequence $\{y^m\} \to y$ and an integer $M$ such that $y^m \in f(x^m), \forall m > M$. In words, this says that for all points in $f(x)$, for

---

1A set is open if, for any point in the set, the $\varepsilon$-neighborhood of that point is also in the set. A set is closed if its complement is open (so closure can depend on which superset the complement is taken relative to).

2Compact has many equivalent definitions, but the most useful is that, in a Euclidean space, compact $\iff$ closed and bounded (by the Heine-Borel theorem).
all sequences in the domain that converge to \( x \), we can come up with a convergent sequence in the range that makes a path that approaches \((x, y)\) from within the graph of the correspondence.

So, upper hemicontinuity states that all convergent paths in the graph of a correspondence have limit points within the graph of the correspondence, while lower hemi-continuity states that for all possible limit points in the graph, there is a convergent path within the graph that approaches it. The two ideas are essentially converses of each other.

It turns out that these definitions are quite different. For a correspondence to be considered **continuous**, it would have to be both uhc and lhc.

Also, note that in the limiting case of functions, uhc and lhc are both equivalent to regular old run-of-the-mill function continuity. So, let’s test our understanding of the definitions a bit. Figure 7.2a) is not uhc, but is lhc. The reason is that the graph is not closed. Contrarily, b) is not lhc, but is uhc. The counterexample sequence is indicated with the small arrow heads.

### 7.1.3 Fixed point theorems

So, now we move in the general direction of understanding the Kakutani fixed point theorem. To do this though, it is instructive to examine Brouwer’s fixed point theorem which deals with functions instead of correspondences.

**Brouwer’s Fixed Point Theorem.** If \( f : A \hookrightarrow A \) is continuous, and \( A \) is compact and convex, then \( f(x) = x \) for some \( x \in A \) (i.e. \( \exists \) a fixed point).

How can we see this? Look at a simple case where \( A = [0, 1] \).
How can we get across \( A \) without crossing the 45° line (i.e. without yielding a fixed point)? With a discontinuity! So if we stop those, we should be OK. Where else might we have trouble? What if \( A = [0, \frac{1}{3}] \cup [\frac{2}{3}, 1] \)?

The fixed point is not in \( A \). Making \( A \) closed and compact simply prevents this sort of counterexample.

So, how might we extend our hypotheses in order for this same sort of idea to fly for correspondences?

First, we need some sort of generalized notion of continuity. In the case of Kakutani’s theorem, upper hemicontinuity fills the void. Is this good enough? Can you think of a case where a uhc correspondence on a convex space does not have a fixed point?
This correspondence is uhc and has no fixed point. How could we remedy this gap in our logic? We could insist that \( f(x) \) is convex \( \forall x \in A \) as well as non empty. Then, we have recaptured the same idea from Brouwer – if we can draw the correspondence without lifting our pen, then it must cross the 45° line, and thus a fixed point must exist. Formally, we have

**Kakutani’s fixed point theorem.** *If a correspondence* \( f : A \rightrightarrows A \) *is upper hemi-continuous and* \( f(x) \) *is convex and non-empty* \( \forall x \in A \), *then* \( f \) *has a fixed point.*

This mathematical background is key for existence proofs of Nash equilibrium, as well as Walrasian equilibrium, as we saw in class.

As a parting shot, why did we not use upper hemicontinuity instead of lower hemicontinuity? Basically, the fact that the graph of a lower hemicontinuous correspondence can be open is hard to overcome. It leaves us susceptible to a counterexample in the vein of the following figure.
7.2 Existence of Walrasian equilibrium

7.2.1 Guiding problem (from a past 202 final)

Consider an economy with 2 agents and 2 commodities (a 2×2 setup). The agents have the following Cobb-Douglas utility functions:

\[ u^1(x, y) = \alpha \log(x) + (1 - \alpha) \log(y) \]  
\[ u^2(x, y) = \beta \log(x) + (1 - \beta) \log(y) \]

with \( \alpha \in (0, 1) \) and \( \beta \in (0, 1) \).

Individual endowments are given by \( e^1 = (0, 1) \) and \( e^2 = (1, 0) \).

a) Derive each agent’s Marshallian demand as a function of \( p_x \) and \( p_y \).

b) Normalize \( p_x = 1 \). Compute the Walrasian equilibrium allocation and prices.

Now suppose a firm enters this market. The firm is a profit maximizer and produces good \( y \) using \( x \) as an input. The firm’s production function is given by \( f(x) = x \), but (unlike what we have studied in class) has a fixed cost of production given by \( F > 0 \). The firm only pays the fixed cost if it produces. If the firm produces nothing, it does not have to pay the fixed cost.

c) Suppose \( \alpha = \frac{2}{5} \) and \( \beta = \frac{4}{5} \). Does a Walrasian equilibrium exist? If so, give the Walrasian equilibrium allocation and prices. If not, explain why an equilibrium does not exist.

d) Suppose \( \alpha = \frac{1}{5} \) and \( \beta = \frac{3}{5} \). Does a Walrasian equilibrium exist? If so, give the Walrasian equilibrium allocation and prices. If not, explain why an equilibrium does not exist.

7.2.2 Notation

In simple consumer theory problems, we used unstarred variables to denote a choice variable or an underdetermined quantity, and we used starred variables to denote an optimum choice. In general equilibrium problems, we have the same distinction to make for prices. Usually though, prices are not chosen by the agents in the general equilibrium model, but rather are quantities that are “chosen” so that maximizers will clear markets. In this case, the prices are not optima in the sense that nobody chooses them. Rather, the “market” chooses them in equilibrium. Hence, we will not denote equilibrium prices with a star, but rather with a hat.

Note that these prices being chosen by the “market” is a bit artificial. This is because we have intentionally generalized away from the complications of price dynamics. All agents have an optimum demand for all goods given the prices. This was the lesson of the first three-quarters of the class. The extra layer that GE adds is that, given the optimum behavior of the agents, we must find prices such
that markets clear. In other words, we are simply looking for resting points of the dynamics. If an omnipotent power were to set the prices, agents would trade at those prices and demand would equal supply. This is the idea of a Walrasian equilibrium.

It is clear that prices which force markets to clear are a natural idea for a market equilibrium, but it is not clear how prices move to those equilibrium quantities. The tâtonnement processes mentioned in the notes were an early attempt to investigate how prices settle to equilibrium\(^3\). The dynamics driving the prices are important because they can tell us about whether a given equilibrium is stable or unstable (or if an equilibrium can even be reached in finite time!). If we nudge prices away from equilibrium, will they return to the equilibrium, or diverge from it? Although these questions are not terribly important for the purposes of this class, they are quite interesting (and for the most part, unanswered).

### 7.2.3 Solution to guiding problem

a) If we exponentiate (an increasing transformation) these utility functions, we see that they are equivalent to a standard Cobb-Douglas utility function. We can then just use the fact that with Cobb-Douglas utility, an agent spends a constant fraction on each good. Hence

\[
\begin{align*}
p_x x_1^* &= \alpha w \\
p_y y_1^* &= (1 - \alpha)w
\end{align*}
\]

Agent 1’s endowment is \(e^1 = (0, 1)\), so his wealth is given by \(w = p_y\). Thus, we find

\[
\begin{align*}
x_1^* &= \frac{\alpha p_y}{p_x} \\
y_1^* &= 1 - \alpha
\end{align*}
\]  

(7.7)

Similarly for agent 2, noting that his endowment is \(e^2 = (1, 0)\), and thus his wealth is \(w = p_x\), we find

\[
\begin{align*}
x_2^* &= \beta \\
y_2^* &= (1 - \beta) \frac{p_x}{p_y}
\end{align*}
\]  

(7.8)

b) In Walrasian equilibrium, markets must clear. We use this to solve for the equilibrium price, \(p_y\).

\[
x_1^* + x_2^* = \alpha p_y + \beta = 1
\]

(7.9)

Does the \(y\) market clear?

\[
y_1^* + y_2^* = 1 - \alpha + \frac{1}{p_y} (1 - \beta) = 1 \Rightarrow p_y = \frac{1 - \beta}{\alpha}
\]

(7.11)

\(^3\)Evolutionary game theory attempts to explain how agents get to Nash equilibrium, while standard game theory is just concerned with what the equilibria are.
So, this was all simple. But in c) and d) we introduce production. What do we do now? Sometimes when a weird production technique is given, the trick is to show that production cannot be used in equilibrium (so that the “with production” equilibrium is the same as the “without production” equilibrium). This is exactly the trick we will use.

Note that production is linear. What does that tell us about profit? Since we have constant returns to scale, \( \pi = 0 \) or \( \pi = +\infty \). The fixed cost doesn’t change this fact. If \( \pi > 0 \) is possible, then \( \pi \to \infty \) is possible. So if the factory can make any profit at all, there can be no equilibrium. But, if the factory makes no profit, then the firm owner would prefer to not open a factory at all and save himself the fixed cost \( F \). So, a profit-maximizing factory owner would choose not to open a factory at all. Hence, in equilibrium, we must somehow force firms to shut down.

c) If \( \alpha = \frac{2}{5} \) and \( \beta = \frac{4}{5} \), then\[
\hat{p}_y = \frac{1/5}{2/5} = \frac{1}{2} \tag{7.12}
\]
This means that \( y \) is cheaper than \( x \). It is thus stupid to use the production technology to transform \( x \) into \( y \). So, the firm will not operate, and the old Walrasian equilibrium is still valid.

d) If \( \alpha = \frac{1}{5} \) and \( \beta = \frac{3}{5} \), then\[
\hat{p}_y = \frac{2/5}{1/5} = 2 \tag{7.13}
\]
Here \( y \) is more expensive, and the firm will demand infinite amounts of \( x \). Hence, no Walrasian equilibrium can exist.

### 7.3 Corner solutions and zero prices in equilibrium problems

#### 7.3.1 Guiding problem (MWG 15.B.9)

Suppose that in a pure exchange economy (i.e., an economy without production), we have two consumers, Alphanse and Betatrix, and two goods, Perrier and Brie. Alphanse and Betatrix have the utility functions:

\[
u_\alpha = \min\{x_{p\alpha}, x_{b\alpha}\} \quad \text{and} \quad u_\beta = \min\{x_{p\beta}, (x_{b\beta})^{1/2}\}
\]

(where \( x_{p\alpha} \) is Alphanse’s consumption of Perrier, and so on). Alphanse starts with an endowment of 30 units of Perrier (and none of Brie); Betatrix starts with 20 units of Brie (and none of Perrier). Neither can consume negative amounts of a good. If the two consumers behave as price takers, what is the equilibrium?

Suppose instead that Alphanse begins with only 5 units of Perrier while Betatrix’s initial endowment remains 20 units of Brie, 0 units of Perrier. What happens now?
7.3.2 Solution to guiding problem

So, for starters, be very worried when you see cute names like “Alphanse” and “Betatrix”. This is a surefire indicator that you are in for a pretty bad problem. If everything has 8 subscripts, chances are the problem is easy and they are just trying to intimidate. So, what does this problem look like? We are in a 2×2 setup, so let’s try to construct an Edgeworth box.

- α will set \( x_p = x_b \) at optimum.
- β will set \( x_p = x_b^{1/2} \) at optimum.

So, we see

\[
\begin{align*}
\alpha &: (x_p, x_b) = (30, x_b) \quad \text{s.t.} \quad x_b \geq 30 \\
\beta &: (x_p, x_b) = (0, x_b) \quad \text{s.t.} \quad x_b \in [0, 20]
\end{align*}
\]

Offer curves never intersect. Does this mean that there is not a Walrasian equilibrium? Certainly not! So where have we gone wrong? When might an optimizing agent not set \( x_p = x_b \) if his utility is \( \min\{x_p, x_b\} \)? With zero prices!

If \( p_b/p_p = 0 \), then β has no income from his endowment. He gets nothing and is indifferent to selling his lousy endowment. Thus, under these prices, his offer curve is

\[
\beta : (x_p, x_b) = (x_p, 20) \quad \text{s.t.} \quad x_p \geq 20^{1/2}
\]

If \( p_p/p_b = 0 \), then α will take at least 20\(^{1/2}\) (≈ 4.8) units of Perrier, so his offer curve becomes

\[
\alpha : (x_p, x_b) = (x_p, 0) \quad \text{s.t.} \quad x_p \in [0, 30]
\]
So, if we have

\[
(x_{p\beta}, x_{b\beta}) = (x_{p\beta}, 20) \\
(x_{p\alpha}, x_{b\alpha}) = (30 - x_{p\beta}, 0)
\] s.t. \(x_{p\beta} \in [20^{1/2}, 30]\) \hspace{1cm} (7.18)

then, we have a Walrasian equilibrium at \(\frac{p_{\alpha}}{p_{b}} = 0\). Where are these equilibria on the Edgeworth box?

So, how will altering endowments change things? First, we look at the updated Edgeworth box.

First, we have a nice interior Walrasian equilibrium. Solving for the intersection\(^4\) of

\[^4\text{Note that this is in terms of } \alpha \text{'s Brie and Perrier. We are actually solving the system}
\]

\[
x_{b\alpha} = x_{p\alpha}^2 \\
x_{b\beta} = (x_{p\beta})^2 \\
x_{b\alpha} + x_{b\beta} = 20 \\
x_{p\alpha} + x_{p\beta} = 5
\]
7.3. CORNER SOLUTIONS AND ZERO PRICES IN EQUILIBRIUM PROBLEMS

\[ y = x \] and \[ y = 20 - (x - 5)^2 \]:

\[
x = 20 - (x - 5)^2 \quad (7.19)
\]

\[
x = 20 - x^2 + 10x - 25
\]

\[ 0 = -x^2 + 9x - 5 \quad (7.20) \]

\[
x = \frac{-9 \pm \sqrt{81 - 20}}{-2} = \frac{-9 \pm \sqrt{61}}{-2} \approx 4.5 \pm 3.9 \quad (7.21)
\]

\[
x^{WE} = \frac{9 - \sqrt{61}}{2} = y^{WE} \quad (7.22)
\]

where the root that is greater than the aggregate supply of 5 has been eliminated.

To get the supporting price, we need the slope of the line from \( \omega \) to \((x^{WE}, y^{WE})\):

\[
p_y = \frac{9 - \sqrt{61}}{2} = \frac{9 - \sqrt{61}}{5 - 9 - \sqrt{61}} = \frac{p_b}{1 + \sqrt{61} p_p} \quad (7.23)
\]

Are we done? What about the boundary cases, when \( p = 0 \) for either good?

If \( p_p/p_b = 0 \), then \( \beta \) will take at least \( 20^{1/2} \approx 4.8 \) units of Perrier, so his offer curve is

\[
\beta : (x_p, x_b) = (x_p, 20) \quad \text{s.t.} \quad x_p \geq 20^{1/2} \quad (7.24)
\]

\( \alpha \) will get no income and is indifferent to selling his worthless Perrier, so

\[
\alpha : (x_p, x_b) = (x_p, 0) \quad \text{s.t.} \quad x_p \in [0, 5] \quad (7.25)
\]

So, if we have

\[
(x_{p\beta}, x_{b\beta}) = (x_{p\beta}, 20) \quad \text{s.t.} \quad x_{p\beta} \in [20^{1/2}, 5] \quad (7.26)
\]

then we have a Walrasian equilibrium with \( p_p/p_b = 0 \).

If \( p_b/p_p = 0 \), then

\[
\beta : (x_p, x_b) = (0, x_b) \quad \text{s.t.} \quad x_b \in [0, 20] \quad (7.27)
\]

\( \alpha \) will get no income and is indifferent to selling his worthless Perrier, so

\[
\alpha : (x_p, x_b) = (5, x_b) \quad \text{s.t.} \quad x_b \geq 5 \quad (7.28)
\]

So, if we have

\[
(x_{p\beta}, x_{b\beta}) = (0, 20 - x_b) \quad \text{s.t.} \quad x_b \in [5, 20] \quad (7.29)
\]

then we have a Walrasian equilibrium with \( p_b/p_p = 0 \). On the Edgeworth box,
7.4 Cutesy tricks with horrible algebra

7.4.1 Guiding problem (MWG 15.B.6)

Compute the equilibria of the following Edgeworth box economy (there is more than one):

\[ u_1(x_{11}, x_{21}) = (x_{11}^{-2} + (12/37)^3 x_{21}^{-2})^{-1/2}, \quad \omega_1 = (1, 0) \]  
(7.30)

\[ u_2(x_{12}, x_{22}) = ((12/37)^3 x_{12}^{-2} + x_{22}^{-2})^{-1/2}, \quad \omega_1 = (0, 1) \]  
(7.31)

(7.32)

7.4.2 Pep talk

Often times, we are given a problem with an impossibly “cute” setup. This sort of thing is obviously directed towards getting a “nice” answer without a calculator. 15.B.6 is obviously one of these. When dealing with seemingly complicated problems, two tricks usually help

(1) Always normalize prices if you can.

(2) Suppress subscripts whenever possible. In solving various parts of the problem, the context usually dictates one or more of the subscripts, so don’t bother writing them down.
7.4.3 Solution to guiding problem

Normalize $p_x = 1$.

Agent 1:

$$\max_{x,y \geq 0} \left[ x^{-2} + \left( \frac{12}{37} \right)^3 y^{-2} \right]^{-1/2} \quad \text{s.t.} \quad py + x \leq 1$$

(7.33)

We are locally non-satiated, so the constraint binds, hence $x = 1 - py$, and we can solve the unconstrained problem. Note that we cannot have the price of either good go to zero because demand is increasing in both arguments, which will cause the demand of the zero-priced good to go to infinity

$$\max_{y \geq 0} \left( (1 - py)^{-2} + \left( \frac{12}{37} \right)^3 y^{-2} \right)^{-1/2}$$

(7.34)

First-order conditions:

$$[y]: -\frac{1}{2} \left[ (1 - py)^{-2} + \left( \frac{12}{37} \right)^3 y^{-2} \right]^{-3/2} \left\{ 2(1 - py)^{-3} p - 2 \left( \frac{12}{37} \right)^3 y^{-3} \right\} = 0$$

(7.35)

The part in square brackets is always non-zero. So, we set the part in curly brackets to zero.

$$(1 - py)^{-3} p = \left( \frac{12}{37} \right)^3 y^{-3}$$

(7.36)

$$py^3 = \left( \frac{12}{37} \right)^3 (1 - py)^3$$

(7.37)

$$p^{1/3} y = \frac{12}{37} (1 - py)$$

(7.38)

$$y = \frac{12}{37} p^{1/3} + \frac{12}{37} p = \frac{1}{\frac{37}{12} p^{1/3} + p} = y$$

(7.39)

$$x = 1 - py = \frac{37}{12} p^{1/3} + p - p = \frac{1}{1 + \frac{12}{37} p^{2/3}} = x$$

(7.40)

Note, that when we took the cube root of both sides of the equation, we chose the branch cut of the multivalued function $y = x^{1/3}$ that assured that Im$(y) = 0$. Thus,

$$\left( x_1^*, y_1^* \right) = \left( \frac{1}{1 + \frac{12}{37} p^{2/3}}, \frac{1}{\frac{37}{12} p^{1/3} + p} \right)$$

(7.41)

OK, let’s buckle down and do the same thing for agent 2. But do we have to? Symmetry is a friend in complicated problems – never neglect to use it! The whole
point of this type of problem is to kill you if you don’t use the tricks. So, if you find
yourself doing an obscene amount of algebra, look for the trick.

What if in agent two’s problem, we make the symmetry substitution \((x, y) \mapsto (y, x)\)? Then, we have

\[
\max_{x, y} \left[ \left( \frac{12}{37} \right) y^{-2} + x^{-2} \right]^{-1/2}
\]

s.t. \(px + y \leq p\) (or \(x + \frac{1}{p} y \leq 1\))

which is just Agent 1’s problem with \(\frac{1}{p}\) instead of \(p\). Hence,

\[
(x^*_2, y^*_2) = \left( \frac{1}{\frac{37}{12} p^{1/3} + \frac{1}{p}}, \frac{1}{\frac{37}{12} p^{-2/3}} \right)
\]

(7.43)

So, we have offer curves. Let’s clear markets. First, keeping track of all these \(\frac{1}{3}\) and \(\frac{2}{3}\) is a pain. Let\(^5\) \(q = p^{1/3}\). Then,

\[
x^*_1 + x^*_2 = \frac{1}{1 + \frac{12}{37} q^2} + \frac{1}{\frac{37}{12} q^{-1} + q^{-3}} = 1
\]

(7.44)

\[
\frac{1}{\frac{37}{12} q^{-1} + q^{-3}} = \frac{\frac{12}{37} q^2}{1 + \frac{12}{37} q^2}
\]

(7.45)

\[
q + \frac{12}{37} q^{-1} = 1 + \frac{12}{37} q^2
\]

(7.46)

Getting rid of negative exponents:

\[
q^2 + \frac{12}{37} q - \frac{12}{37} q^3 = 0
\]

(7.47)

\[
37 q^2 + 12 - 37 q - 12 q^3 = 0
\]

(7.48)

What now? Factoring a cubic is way too hard. What is the trick here? These sorts
of problems almost always have simple answers. That is why they are to ridiculously
fine tuned to start with \( \left( \frac{12}{37} \right)^3 \)?!? So, guess what you would want the answer to
be. How about \(q = 1\)? Plugging it in reveals that it is indeed a root. Now how to
we factor out the \(q - 1\) factor? With synthetic division (which is a useful trick to at
least know how to do if necessary).

\[
\begin{array}{r|rrrrrr}
& -12q^2 & + 25q & - 12 \\
q - 1 & -12q^3 & + 37q^2 & - 37q & + 12 \\
\hline
& 12q^3 & - 12q^2 \\
\end{array}
\]

\[
\begin{array}{r|rrrrrr}
& 25q^2 & - 37q \\
& -25q^2 & + 25q \\
\hline
& -12q & + 12 \\
& 12q & - 12 \\
& 0
\end{array}
\]

\(5\)Relabelling to avoid algebra is another important trick
OK, so we are left with

\[-12q^2 + 25q - 12 = 0\]  \hspace{1cm} (7.50)

So, now we can factor the quadratic. What numbers multiply to 1 but also add to \(-\frac{25}{12}\)? Reciprocals multiply to 1, so consider \(\frac{a}{b}\) and \(\frac{b}{a}\). Their sum is

\[\frac{a^2 + b^2}{ab} = \frac{25}{12}\]  \hspace{1cm} (7.51)

So, what factors of 12, when squared and summed, give 25? We should expect integers. It’s our old friend, the Pythagorean triple \(3^2 + 4^2 = 5^2\). Hence, the complete factoring is

\[12 (q - 1) \left(q - \frac{4}{3}\right) \left(q - \frac{3}{4}\right) = 0\]  \hspace{1cm} (7.52)

Hence, \(q \in \{1, \frac{4}{3}, \frac{3}{4}\}\), or

\[\left\{1, \left(\frac{4}{3}\right)^{1/3}, \left(\frac{3}{4}\right)^{1/3}\right\}\]  \hspace{1cm} (7.53)

What about the market for \(y\)? Do we check to see if it clears? Again, symmetry comes to the rescue. \(y_1^* + y_2^* = 0\) is the same as \(x_1^* + x_2^* = 0\) under the transformation \(p \mapsto \frac{1}{p}\). But, our solutions to \(x_1^* + x_2^* = 0\) are closed under this mapping, so the \(y\)-market clears by symmetry. Done!
Chapter 8

General Equilibrium II

8.1 Continuous GE

8.1.1 Guiding question (#3, June 2005 Comp)

Consider an economy consisting of a continuum of agents of unit mass with abilities distributed uniformly on $[0,1]$. Each agent is endowed with one unit of labor and wants to maximize consumption. There is also a unit mass of firms indexed by $j \in [0,1]$. Each firm has the same technology to transform labor into a single consumption good. The technology requires two workers, one who is assigned a “managerial” task and one a “worker” task. If the manager’s ability is $a_M$ and the worker’s ability is $a_W$, firm output is $y = a_W + a_M + \lambda a_M (1 - a_W)$ where $0 < \lambda < 1$.

(a) Suppose a firm employs two agents with abilities $a$ and $a'$ where $a' > a$. How should it assign the agents in order to produce efficiently?

An allocation specifies hiring by firms and consumptions by workers. Prices are given by a wage schedule $w(a) : [0,1] \mapsto \mathbb{R}_+$ and a price for the consumption good that we can normalize to one.

(b) Define a Walrasian equilibrium in this economy. (NB: You can do this either by formally defining an allocation and conditions for WE, or by giving a clear verbal description.)

(c) Prove that the equilibrium allocation of workers to firms is “negatively assortative” – if a firm hires agent $a$, it also hires agent $1 - a$.

(d) Solve for the Walrasian equilibrium wage schedule.

Parts (e) and (f) ask you to describe how changes in the distribution of worker ability affect Walrasian equilibrium output and wages.

(e) Suppose the ability distribution changes from $U[0,1]$ to a distribution $H$ on $[0,1]$, also symmetric around $\frac{1}{2}$, but a mean preserving spread of the uniform distribution. How will this affect aggregate output?
(f) Suppose the ability distribution changes from $U[0,1]$ to a distribution $H$ on $[0,1]$ where $H(a) = 0$ for $a \in [0, \frac{1}{4}]$, $H(a) = 2a - \frac{1}{2}$ for $a \in \left[\frac{1}{4}, \frac{1}{2}\right]$ and $H(a) = a$ on $[\frac{1}{2}, 1]$. How will this affect the equilibrium wage distribution?

8.1.2 Solution to guiding question

(a) 

\begin{align*}
y &= a_W + a_M + \lambda a_M - \lambda a_M a_W \\
y &= a_W + (1 + \lambda)a_M - \lambda a_M a_W \\
\end{align*}

$\lambda a_M$ has the higher return, so we clearly choose the higher $\lambda$ to be manager.

(b) A Walrasian is an allocation and prices such that everyone optimizes and markets clear.

How can we write this mathematically? Well, all agents have an endowment of labor that is worth no utility to them. So, since the wage schedule is positive, they will all work. Hence, full employment is the equivalent of agents optimizing.

Firms optimizing means that they choose to employ an $(a_W, a_M)$ pair that is an element of $\arg\max_{(a,a')}(y(a,a') - w(a) - w(a'))$ and that the value of the objective at this pair is weakly greater than zero (so that the firm does not prefer to shut down). Each firm can only hire 0 agents (shut down) or 2.

Finally, we have two markets to clear. Labor market clearing is best stated in terms of sets and their measure (given by $\mu$). If a set of firms $F$ hires a set of managers $M$ and a set of workers $W$, then we must have that $\mu(F) = \mu(M) = \mu(W)$ since there is a one-to-one correspondence with operating firms and manager-agent pairs.

Product market clearing means that all of the consumption good produced by the firm must be sold to the agents in exchange for labor. In other words, the firms employs agents, the agents plug their labor endowment into the firm’s production function, and they are paid in the consumption good produced by the firm. If wages don’t equal output, then there is leftover production good that no one can buy (we assume here that agents only have one employment opportunity – working for the manager-agent firms described above). Hence we must have that firms make zero-profit to clear the product market. Another way to see this is that there is a unit mass of firms and a unit mass of agents. Each firm must hire two agents, so only half of the firms will be operating if there is full employment. Hence, since the firms are homogeneous, they must be indifferent between shutting down and producing. So, firms must make zero profit.

In sum, the conditions for Walrasian equilibrium are

- All agents are employed (agents optimize).
• Firms make zero profit (product market clears).
• If a set of operating firms $F$ employs a set of managers $M$ and a set of workers $W$, then $\mu(F) = \mu(M) = \mu(W)$ (employers hire two workers or none).
• Operating firms must employ worker-manager pairs that are in the arg max and that produce a non-negative profit (firms optimize).

\begin{equation}
\frac{\partial^2 \Pi}{\partial a_M \partial a_W} = -\lambda < 0 \tag{8.3}
\end{equation}

So, our profit function is submodular in $(a_M, a_W)$. Topkis then requires that $a_M^*(a_W)$ is non-increasing. So a firm that chooses a higher $a_M$ must choose a lower $a_W$. This is like saying that managers and workers act as substitutes and not as complements. Given this idea, here are two ways to reach the assortative conclusion

(i) The intuitive way

How does this substitutes condition help us? Say that we have an infinitessimal chunk of the firm continuum, $dj$. This chunk of the firm continuum must hire a managerial chunk $da_M$ and a worker chunk $da_W$ from the agent continuum. What’s more, since a firm hires two workers, the measures$^1$ of these chunks must obey $|dj| = |da_W| = |da_M|$, where, $|da_W| = |da_M|$ since a firm must hire the same measure of workers and managers (namely one).$^2$

Now, since we are told the wage function cannot be negative (the agents have limited liability), we know that all agents must work. Consider the set of agents $A = [1 - da, 1]$ where $|A| = \varepsilon$. As the we just illustrated, the measure of the set of firms that hires $A$ must also be $\varepsilon$. And since this is the set of highest ability agents, we have that the agents in $A$ must all be employed as managers. So where is the set of their corresponding workers, $B$. Clearly, by the Topkis results above, all $a < \min\{x|x \in B\}$ must remain unemployed, since there is no one left of higher ability to hire them than those in set $A$. Thus, to ensure that all workers are hired, we must have that $\min\{x|x \in B\} = 0$, and since the measure on the agent continuum is uniform, we have that $B = [0, da]$. Now we can repeat the same process with the superset $[da, 1 - da]$.

If we continue with chunks of measure $\varepsilon$, it is clear that we will continue the same process until the managerial and workers chunks run into each other at $a = \frac{1}{2}$. As $\varepsilon \to 0$, it is clear that we are simply matching workers and managers by moving up from 0 and down from 1 linearly with slope 1. The only way to make these who trends meet up in the middle is then to have $a_W = 1 - a_M$. Hence matching is assortative.

$^1$The measure of a set $A$ under a probability distribution with cdf $F(a)$ is just $|A| = \int_{a \in A} dF(a)$.

$^2$When doing anything that uses measure theory, a good rule of thumb is that the word “number” is replaced with the word “measure”.
(ii) **The mathematical way**

Say that the assortment scheme matches managers with workers via the function $g$, i.e. $a_M = g(a_W)$. If we assume this mapping is continuous, then a chunk of worker abilities $[a_W, a_W + da_W]$ maps to a chunk of manager abilities $[a_M + da_M, a_M]$, where $da_M$ is related to $da_W$ by

$$
da_M = g'(a_W)da_W$$

(8.4)

Note that $da_M$ is negative, since $g$ is a decreasing function. Now, since these two ability chunks will be matched to each other, we must have that they are of the same measure. So, if $f$ is the PDF of the distribution of abilities, then

$$|da_M|f(a_M) = |da_W|f(a_W)$$

$$-g'(a_W) \cdot 1 = 1$$

(8.5)

So, the assortment scheme $g$ solves the differential equation $g' = -1$. The general solution to this is $g(a) = -a + C$ where the constant $C$ is determined by initial conditions. We want to have the lowest ability worker matched with the highest ability manager, so we set $g(0) = C = 1$. So our final assortment scheme is given by

$$g(a) = 1 - a$$

(8.6)

(d) Firms must solve their maximization problem. Its first-order conditions are

$$[a_M] : 1 + \lambda (1 - a_W) = w'(a_M)$$

$$[a_W] : 1 - \lambda a_M = w'(a_W)$$

(8.7)

Due to our assortative conclusion, we must find that $a_M \in \left[\frac{1}{2}, 1\right]$ and $a_W \in \left[0, \frac{1}{2}\right]$ in equilibrium. So, our first-order conditions provide a first-order differential equation for the wage schedule

$$w'(a) = \begin{cases} 
1 - \lambda a_M & a \in \left[0, \frac{1}{2}\right] \\
1 + \lambda (1 - a_W) & a \in \left[\frac{1}{2}, 1\right]
\end{cases}$$

(8.8)

Integrating, we find

$$w(a) = \begin{cases} 
(1 - \lambda)a + \frac{1}{2}\lambda a^2 + A & a \in \left[0, \frac{1}{2}\right] \\
a + \frac{1}{2}\lambda a^2 + B & a \in \left[\frac{1}{2}, 1\right]
\end{cases}$$

(8.9)

---

3If we have two managers with abilities $m'$ and $m$, $m' > m$, then our Topkis logic shows us that their matched worker abilities must be oppositely oriented, that is $w' < w$. So, if the highest ability managers are not matched with the lowest ability workers, then no one can match with the lowest ability workers and there will be unemployment in equilibrium. Hence the highest managers match with the lowest workers.
where $A$ and $B$ are constants of integration. How do we determine $A$ and $B$?

First, it must be continuous. At a discontinuity, firms would do better to make a small shift in $a$ to get the discount afforded by the discontinuity. Agents on the more expensive side of the discontinuity would thus be unemployed, which cannot be in equilibrium. So, setting $a = \frac{1}{2}$, we solve

$$
(I - \lambda)\frac{1}{2} + \frac{1}{8}\lambda + A = \frac{1}{2} + \frac{1}{8}\lambda + B \quad (8.10)
$$

Can we further restrict $A$ and $B$?

Note that we have a unit mass of firms. Each firm hires two agents; however, there is only a unit mass of them. So, half the firms must choose not to produce. Hence, firm profit must be zero in equilibrium.

$$
\Pi = a_W + a_M(1 + \lambda) - \lambda a_M a_M - w(a_M) - w(a_W) = 0 \quad (8.12)
$$

$$
0 = 1 - a_M + a_M + \lambda a_M - \lambda a_M + \lambda a_M^2 - a_M - \frac{1}{2} \lambda a_M^2 - \frac{B}{a - \frac{1}{4} \lambda} \quad (8.13)
$$

$$
- (1 - \lambda)(1 - a_M) - \frac{1}{2} \lambda (1 - a_M)^2 - A
$$

After a great deal of algebra, we find $A = \frac{1}{2}$ and $B = 0$.

$$
w(a) = \begin{cases} 
\frac{1}{2} \lambda a^2 + (1 - \lambda)a + \frac{1}{2} \lambda & a \in [0, \frac{1}{2}] \\
\frac{1}{2} \lambda a^2 & a \in [\frac{1}{2}, 1] 
\end{cases} \quad (8.14)
$$

(e) So, what does a firm that hires $a_M = a$ and $a_W = 1 - a$ produce?

$$
y = a_M + a_W + \lambda a_M (1 - a_W) = a + 1 - a + \lambda a(1 - (1 - a)) = 1 + \lambda a^2 \quad (8.15)
$$

So, higher $a$ leads to more output. Note that with a mean preserving spread that is symmetric about $\frac{1}{2}$, the logic of (c) still holds – we are still negatively assortative in equilibrium (see the note on assortment and continuous general equilibrium for a better look at the math that explains why this is true). Our new distribution simply puts more weight on managers with higher $a$ values (and workers with lower $a$ values). Another way to think about this is that the symmetry means you are increasing manager ability and decreasing worker ability by the same amount for each worker-manager pair. But managers have a higher return (as we saw in (a)), so production goes up. **So, output will go up under the new distribution.**

(f) What has happened here? We have pretty much just squished all of the agents that were on $a \in [0, \frac{1}{2}]$ into the smaller interval $a \in [\frac{1}{4}, \frac{1}{2}]$. Now, again we consider an intuitive and a mathematical way to solve for the new assortment scheme.
(i) **The intuitive way**

So, to make things work, we are still negatively assortative, but instead of \( a_W = 1 - a_M \), we must slow the rate of “intrusion” into \([\frac{1}{4}, \frac{1}{2}]\) by a factor of two. So, we consider an assortion scheme of form \( a_W = \xi - \frac{1}{2}a_M \).

What then is \( \xi \)?

We still need \( a_M = a_W = \frac{1}{2} \) to be part of the scheme, by our logic in (c). So,

\[
a_W = \frac{1}{2} = \xi - \frac{1}{2} \cdot \frac{1}{2} = \xi - \frac{1}{2}a_M \quad (8.16)
\]

\[
\xi = \frac{3}{4} \quad (8.17)
\]

So, our assortion scheme is \( a_W = \frac{3}{4} - \frac{1}{2}a_M \) (or \( a_M = \frac{3}{2} - 2a_W \)).

(ii) We can just go back to the same ideas we used in part (c). Then,

\[
\frac{|da_M|f(a_M)}{-g'(a_W) \cdot 1} = 2 \quad (8.18)
\]

Hence the assortion scheme is the solution to \( g' = -2 \), which has general solution \( g(a) = -2a + C \). We match the 1/4 worker with the 1 manager, so we must have \( g(1/4) = -1/2 + C = 1 \). So, \( C = 3/2 \) and our assortion scheme is

\[
a_M = g(a_W) = \frac{3}{2} - 2a_W \quad (8.19)
\]

or

\[
a_W = g^{-1}(a_M) = \frac{3}{4} - \frac{1}{2}a_M \quad (8.20)
\]

Our first-order conditions are still the same

\[
\begin{align*}
[a_M] : & \quad 1 + \lambda(1 - a_W) = w'(a_M) \\
[a_W] : & \quad 1 - \lambda a_M = w'(a_W)
\end{align*} \quad (8.21)
\]

But, we must plug a different assortion scheme into them. Doing so yields

\[
w'(a) = \begin{cases} 
1 - \lambda(\frac{3}{2} - 2a) & a \in [\frac{1}{4}, \frac{1}{2}] \\
1 + \lambda(\frac{1}{4} + \frac{1}{2}a) & a \in [\frac{1}{2}, 1]
\end{cases} \quad (8.22)
\]

Integrating

\[
w(a) = \begin{cases} 
(1 - \frac{3}{2}\lambda)a + \lambda a^2 + C & a \in [\frac{1}{4}, \frac{1}{2}] \\
(1 + \frac{1}{4}\lambda)a + \frac{1}{4}\lambda a^2 + D & a \in [\frac{1}{2}, 1]
\end{cases} \quad (8.23)
\]

To assure continuity, we need \( C = D + \frac{11}{16}\lambda \).

And again, we must have \( \Pi = 0 \). Solving this gives us

\[
C = \frac{5}{8}\lambda \quad D = -\frac{1}{16}\lambda \quad (8.24)
\]

So,

\[
w(a) = \begin{cases} 
(1 - \frac{3}{2}\lambda)a + \lambda a^2 + \frac{5}{8}\lambda & a \in [\frac{1}{4}, \frac{1}{2}] \\
(1 + \frac{1}{4}\lambda)a + \frac{1}{4}\lambda a^2 - \frac{1}{16}\lambda & a \in [\frac{1}{2}, 1]
\end{cases} \quad (8.25)
\]
We are comparing this to the wage schedule under the old $U[0, 1]$ agent distribution

$$w_{OLD}(a) = \begin{cases} \frac{1}{2} \lambda a^2 + (1 - \lambda)a + \frac{1}{2} \lambda & a \in [0, \frac{1}{2}] \\ a + \frac{1}{2} \lambda a^2 & a \in [\frac{1}{2}, 1] \end{cases} \quad (8.26)$$

So, when $a < \frac{1}{2}$, we compare

$$w - w_{OLD} = \lambda \left( -\frac{1}{2}a + a^2 + \frac{1}{8} \right) > 0 \quad \forall a \in \left[ \frac{1}{2}, \frac{1}{2} \right] \quad (8.27)$$

When $a > \frac{1}{2}$

$$w - w_{OLD} = \lambda \left( \frac{1}{4}a - a^2 - \frac{1}{16} \right) < 0 \quad \forall a \in \left[ \frac{1}{2}, 1 \right] \quad (8.28)$$

So, the workers get paid more under the new distribution and the managers get paid less. This makes a lot of sense. Since there are more higher level workers, the marginal product of a manager decreases, since $\frac{\partial^2 \Pi}{\partial a \partial a_M} < 0$. For completeness’ sake, here is a plot of $w - w_{OLD}$ for $\lambda = 0.6$.

![Plot of w - w_{OLD} for lambda = 0.6](image.png)

### 8.2 Market failure due to externality

#### 8.2.1 Guiding question (#3, June 2004 Comp)

Consider a country with two tradeable commodities, plastic and oil, and a third commodity, pollution, that is not priced or traded. There is a single firm with a constant returns technology that converts one unit of oil into one unit of plastic and one unit of pollution. A production plan $y = (y_1, y_2, y_3)$ specifies production of plastic, oil, and pollution. The firm’s production possibility set is $Y = \{(\gamma, -\gamma, \gamma) : \gamma \geq 0\}$. There is a single consumer with utility

$$u(x_1, x_2, y_3) = \log x_1 + \log x_2 - \frac{3}{2}y_3 \quad (8.29)$$

where $x_1, x_2$ are her consumption of plastic and oil, and $y_3$ is the firm’s production of pollution. Suppose the consumer is endowed with one unit of oil. Let $p_1, p_2 \geq 0$ denote prices for plastic and oil, and normalize $p_2 = 1$ so oil is the numéraire.
(a) Solve for the Walrasian equilibrium prices and quantities in this economy.

(b) Solve for the Pareto efficient allocation and compare it to the WE.

(c) Can the government decentralize the Pareto efficient allocation by choosing a suitable tax on plastic and redistributing the proceeds to the consumer as a lump-sum transfer?

8.2.2 Externality

In the notes, we proved the welfare theorems, but under the assumption that there were no externalities. One of the most common wedges between the social optimum and the market outcome is externality. Here, this fact will be illustrated in the setting of general equilibrium with production.

8.2.3 Solution to guiding problem

(a) All ordered triples are of form \((\text{plastic}, \text{oil}, \text{pollution})\).

The firm must solve

\[
\max_{y_2 \geq 0} (p_1 - p_2)y_2
\]

So, for an equilibrium to exist, the firm must make zero profit or shut down, hence, we must have \(\hat{p}_1 \leq \hat{p}_2\). But, since the consumer has infinite marginal utility of plastic with her endowment, she will demand plastic at any price. So, we must have production in equilibrium. Hence \(\hat{p}_1 = \hat{p}_2\).

The consumer solves

\[
\max_{x_1, x_2} \log x_1 + \log x_2 - \frac{3}{2}y_3
\]

\[\text{s.t. } p_1 x_1 + p_2 x_2 \leq p_2 \]

(Note that \(y_3\) is not a choice variable – the agent treats this as an externality that she cannot change. This externality will be the wedge that separates the Walrasian equilibrium from the Pareto optimum.)

The budget constraint binds (local nonsatiation), thus \(x_2 = \frac{1}{p_2} (p_2 - p_1 x_1) = 1 - \frac{p_1}{p_2} x_1\).

\[
\max_{x_1} \log x_1 + \log(p_2 - p_1 x) - \log p_2 - \frac{3}{2}y_3
\]

\[\text{s.t. } x_1 \]

\[\begin{bmatrix} x_1 \end{bmatrix} : \frac{1}{x_1} = \frac{p_1}{p_2 - p_1 x_1} \Rightarrow p_1 x_1 = p_2 - p_1 x_1 \Rightarrow x_1^* = \frac{1}{2} \frac{p_2}{p_1} \Rightarrow x_2^* = \frac{1}{2} \]

So, our Walrasian equilibrium is given by

\[
\hat{p}_1 = \hat{p}_2
\]

\[
x_1^* = x_2^* = \frac{1}{2}
\]

\[
y^* = \left(\frac{1}{2}, -\frac{1}{2}, \frac{1}{2}\right)
\]
(b) The all-knowing social planner sees the externality and controls the firm only for the good of society. He chooses the amount of plastic, $\gamma$, he should produce to make the consumer best off

$$\max_{0 \leq \gamma \leq 1} \log \gamma + \log(1 - \gamma) - \frac{3}{2} \gamma$$  \hspace{1cm} (8.37)

$$[\gamma] : \frac{1}{\gamma} - \frac{3}{2} = \frac{1}{1 - \gamma} \Rightarrow 2 - 3\gamma = \frac{1}{1 - \gamma} \Rightarrow \begin{cases} 2\gamma = (2 - 3\gamma)(1 - \gamma) \\ 2\gamma = 2 - 3\gamma - 2\gamma + 3\gamma^2 \\ 0 = 3\gamma^2 - 7\gamma + 2 \\ \gamma = \frac{7 \pm \sqrt{49-24}}{6} \\ \gamma = \frac{7 \pm 5}{6} = \{\frac{1}{3}, 2\} \end{cases}$$  \hspace{1cm} (8.38)

$$\gamma^* = \frac{1}{3}$$  \hspace{1cm} (8.39)

(where we eliminated the root 2 as extraneous, since $\log(-1)$ is imaginary$^4$)

So,

$$x^{PO}_1 = \frac{1}{3}, \quad y^{PO} = \left(\frac{1}{3} - \frac{1}{3}, \frac{1}{3}\right)$$

is the Pareto optimal allocation.  \hspace{1cm} (8.40)

(c) So, the firm’s production technology necessitates $\hat{p}_1 = \hat{p}_2$ in Walrasian equilibrium. But, we need the effective price of plastic to the consumer to be higher than that of oil in order to reflect the negative externality that pollution inflicts upon her. So, we apply a proportional tax to plastic spending. This tax is then refunded to consumers in the form of a non-distortionary lump-sum rebate.

So, the agent solves

$$\max_{x_1, x_2} \log x_1 + \log x_2$$

s.t.

$$\underbrace{(1 + \tau)}_{\text{proportional tax rate}} \quad p_1 x_1 + p_2 x_2 \leq p_2 + \lambda$$

$$\underbrace{\text{lump-sum rebate}}_{\text{tax revenue}}$$  \hspace{1cm} (8.41)

The constraint binds, so $x_2 = \frac{1}{p_2}[p_2 + \lambda - (1 + \tau)p_1 x_1]$. So,

$$\max_{x_1} \log x_1 + \log[p_2 + \lambda - (1 + \tau)p_1 x_1]$$  \hspace{1cm} (8.42)

$$[x_1] : \frac{1}{x_1} = \frac{(1 + \tau)p_1}{p_2 + \lambda - (1 + \tau)p_1 x_1} \Rightarrow p_2 + \lambda - (1 + \tau)p_1 x_1 = (1 + \tau)p_1 x_1$$  \hspace{1cm} (8.43)

What is $\lambda$? Now that we have calculated the first-order conditions, we can plug it into them$^5$. It is the revenue from the tax, namely $\lambda = \tau p_1 x_1$.

$$p_2 + \tau p_1 x_1 - p_1 x_1 + \tau p_1 x_1 = p_1 x_1 + \tau p_1 x_1$$

$$p_2 = (2p_1 - \tau p_1)x_1$$  \hspace{1cm} (8.44)

$$x_1^* = \frac{p_2}{p_1(2 + \tau)}$$  \hspace{1cm} (8.45)

---

$^4$\[e^{\pi i} = \cos(\pi) + i \sin(\pi) = -1 + 0 \cdot i = -1. \text{ Hence, } \log(-1) = \pi i.\]

$^5$Note that $\lambda$ contains the choice variable. If we had plugged into the objective function before taking first-order conditions, we would have been solving the problem where the consumer knows
We know that zero profit and necessity of production require that \( \hat{p}_1 = \hat{p}_2 \), so, if we set \( \tau = 1 \), we get \( x_1^* = \frac{1}{3} = x_1^{PO} \). What about \( x_2^* \)?

\[
x_2^* = \frac{1}{p_2} [p_2 + \tau p_1 x_1 - p_1 x_1 - \tau p_1 x_1] = \frac{1}{p_2} [p_2 - p_1 x_1] = 1 - \frac{p_1}{p_2} x_1^* = 1 - \frac{1}{2 + \tau}
\]  

(8.47)

So, evaluating at \( \tau = 1 \),

\[
x_2^* \big|_{\tau=1} = \frac{2}{3} = x_2^{PO}
\]

(8.48)

So if we place a tax of \( \tau = 1 \) on plastic spending and then refund the tax revenue as a lump-sum rebate, then the decentralized Walrasian equilibrium will also be the Pareto optimal allocation.

As a parting shot, note that we could have just as easily used a tax on the producer’s use of oil (and thus his pollution) to implement the Pareto optimal outcome in equilibrium. Either approach is acceptable.

8.3 Market failure due to overproduction

8.3.1 Guiding question (MWG 16.G.5)

In a certain economy there are two commodities, education \( e \) and food \( f \), produced by using labor \( L \) and land \( T \) according to the production functions

\[
e = (\min\{L, T\})^2 \quad \text{and} \quad f = (LT)^{1/2}
\]

(8.49)

There is a single consumer with the utility function

\[
u(e, f) = e^\alpha f^{1-\alpha}
\]

(8.50)

and endowment \((\omega_L, \omega_T)\). To ease the calculations, take \( \omega_L = \omega_T = 1 \) and \( \alpha = \frac{1}{2} \).

(a) Find the optimal allocation of the endowments to their productive uses.

(b) Recognizing that the production for education entails increasing returns to scale, the government of this economy decides to control the education industry and finance its operation with a lump-sum tax on the consumer. The consumption of education is competitive in the sense that the consumer can choose any amount of education and food desired at the going prices. The food industry remains competitive in both its production and consumption aspects. Assuming that the education industry minimizes cost, find the marginal cost of education at the optimum.

that she controls the lump sum rebate. But, we are trying to model a world where she is one of many taxpayers and has no control over her rebate. To do this, we force her to make a decision as if \( \lambda \) were an externality. Only after she has made her choice (i.e. we have derived the first-order conditions) can we plug in what \( \lambda \) actually is. This is the same idea as when you solve for a symmetric Cournot equilibrium. You can’t plug in the fact that the other sellers also sell \( q \) until after you have taken the first-order conditions assuming that other sellers’ quantities are exogenous.
Show that if this price of education were announced, together with a lump-sum tax to finance the deficit incurred when the education sector produces the optimal amount at this price for its product, then the consumer’s choice of education will be at the optimal level.

(c) What is the level of the lump-sum tax necessary to decentralize this optimum in the manner described in (b)?

Now suppose there are two consumers and that their preferences are identical to those above. One owns all of the land and the other owns all of the labor. In this society, arbitrary lump-sum taxes are not possible. It is the law that any deficit incurred by a public enterprise must be covered by a tax on the value of land.

(d) In the appropriate notation, write the transfer from the landowner as a function of the government’s planned production of education.

(e) Find a marginal cost price equilibrium for this economy where transfers have to be compatible with the transfer function specified in (d). Is it Pareto optimal?

8.3.2 Overproduction

In the notes, the proof of the welfare theorems with production rests on the assumption that the firm’s technology is not increasing returns to scale. If it is, the resulting overproduction drives a wedge between the social optimum and the decentralized Walrasian equilibrium.

8.3.3 Solution to guiding problem

(a) So, this is the social planner’s problem. Say we devote \( L_e \) of the labor endowment to education and \( T_e \) of the land endowment. We will exhaust our budget, so the land and labor devoted to food are given by

\[
L_f = 1 - L_e \\
T_f = 1 - T_e
\]

Hence, our original social planner’s problem,

\[
\max_{L_e,L_f,T_e,T_f \geq 0} u[e(L_e,T_e), f(L_f,T_f)] \\
\text{s.t.} \quad L_e + L_f \leq 1 \\
T_e + T_f \leq 1
\]

reduces to

\[
\max_{0 \leq L_e, T_e \leq 1} u[e(L_e,T_e), f(1 - L_e, 1 - T_e)]
\]

or more specifically,

\[
\max_{0 \leq L_e, T_e \leq 1} \min \{L_e, T_e\} \ (1 - L_e)^{1/4} (1 - T_e)^{1/4}
\]
The min immediately suggests that $L_e = T_e$. Is there any reason that this might be suboptimal? Say $L_e > T_e$. Then by reducing $L_e$, we increase the $(1 - L_e)^{1/4}$ term while leaving the min unchanged. The same logic holds for $T_e > L_e$, so we conclude that $L_e = T_e$, yielding

$$\max_{0 \leq L_e \leq 1} L_e (1 - L_e)^{1/2}$$

(8.56)

First-order conditions then require that

$$[L_e] : \begin{cases} (1 - L_e)^{1/2} = \frac{1}{2} L_e (1 - L_e)^{-1/2} \\ 1 - L_e = \frac{1}{2} L_e \\ L_e = \frac{2}{3} \end{cases}$$

(8.57)

So, the social optimum is $L_e^* = T_e^* = \frac{2}{3}$, $L_f^*, T_f^* = \frac{1}{3}$ which leads to $e^* = \frac{4}{9}$, $f^* = \frac{1}{3}$.

(b, c) Now the phrasing of this problem is a bit obscure. What is the big idea? We just showed what the social optimum is. Will this optimum be reached if we just leave society to compete in the market? Or, more succinctly, can this optimum be supported by prices in a competitive market?

The quick answer is that it cannot. The education production function has increasing returns to scale, so if producing any education level $e$ is profitable, it is even more profitable to produce $e' > e$. Hence, if profit is possible, the profit maximizing education industry will produce $e = 1$, which clearly suboptimal for society. There is no way to competitively make it produce $e^* = \frac{4}{9}$ because any time that $e = \frac{4}{9}$ is preferred to shutdown, $e = 1$ is preferred to $e = \frac{4}{9}$. So, how do we fix this?

We have the government run the education industry. The government’s objective is then to produce the optimal $e^* = \frac{4}{9}$ at the lowest possible cost. But, since the government is necessarily unconcerned with profit, it could run a deficit in producing $e^* = \frac{4}{9}$. How do we then fund this deficit (we need a balanced government budget)? With non-distortionary (lump-sum) taxes. OK, so now we have the “big idea” in mind. Can we achieve the optimal outcome through government takeover of the “problem” industry – in this case, education?

Government minimizes cost to produce $e^* = \frac{4}{9}$. So $T_e = L_e = \frac{2}{3}$ necessarily. Can we support this? What about the food industry? It must solve

$$\max_{L_f, T_f \geq 0} p_f (L_f T_f)^{1/2} - p_L L_f - p_T T_f$$

(8.58)

$$[L_f] : \begin{cases} \frac{1}{2} p_f L_f^{1/2} T_f^{1/2} = p_L \\ [T_f] : \begin{cases} \frac{1}{2} p_f L_f^{1/2} T_f^{1/2} = p_T \end{cases}$$

(8.59)

---

6The allocation will be interior. For $L_e < 0$, the objective is negative, while for $L_e > 0$ it is positive, and for $L_e > 1$, the objective become imaginary, which is not permissible (the utility function’s domain is $L_e \leq 1$).
Thus, \([L_f] \cdot [T_f]\) tells us that \(p_f = 2\sqrt{p_L p_T}\) and \([L_f]/[T_f]\) tells us \(\frac{T_f}{L_f} = \frac{p_L}{p_T}\). So government efficiency tells us that \(\frac{T_f}{L_f} = 1 - \frac{T_e}{L_f} = 1\). As a result, food industry profit maximization requires that \(p_L = p_T = \frac{1}{2}p_f\). Hence, we normalize \(p_L = 1\) yielding

\[
\begin{align*}
    p_L &= p_T = 1 \\
    p_f &= 2
\end{align*}
\]

So, can we find a price for education, \(p_e\) to support it all? First of all, what is the surplus/deficit that the government has in running the education industry optimally for society.

\[
\Pi = p_e e^* - p_L L_e^* - p_T T_e^*
\]

\[
\Pi = \frac{4}{9}p_e - \frac{2}{3} - \frac{4}{3} = \frac{4}{9}p_e - \frac{4}{3}
\]

So, the government distributes this surplus/deficit in the form of a lump sum tax, \(\tau = \Pi\). The consumer solves

\[
\max_{e,f} \quad e^{1/2}f^{1/2}
\quad \text{s.t.} \quad \frac{3}{2}e + 2f \leq 2 - \tau = 2 - \Pi
\]

The budget constraint binds, so \(f = \frac{1}{2}(\frac{2}{3} + \frac{4}{9}p_e - p_e e)\). Hence, the simplified CP is

\[
\max_e e^{1/2} \left( \frac{1}{2} \left( \frac{2}{3} + \frac{4}{9}p_e - p_e e \right) \right)^{1/2}
\]

\[
[e] : \begin{cases}
    \frac{4}{9}e^{* - 1/2}(\left( \frac{2}{3} + \frac{4}{9}p_e - p_e e^* \right)^{1/2} = \frac{4}{9}e^{1/2}p_e(\left( \frac{2}{3} + \frac{4}{9}p_e - p_e e^* \right))^{-1/2} \\
    \frac{2}{3} + \frac{4}{9}p_e - p_e e^* = p_e e^* \\
    \frac{1}{2}p_e(\frac{2}{3} + \frac{4}{9}p_e) = e^*
\end{cases}
\]

so, if we want the demand for education to clear the market, we must have \(e^* = \frac{4}{9} = e^{PO}\). Solving for the price of education that implements this,

\[
\frac{2}{3} + \frac{4}{9}p_e = \frac{8}{9}p_e
\]

\[
\frac{2}{3} = \frac{4}{9}p_e
\]

\[
\frac{3}{2} = p_e
\]

so, \(p_e = \frac{3}{2}\) makes the consumer demand the optimum level of education. This price is quite interesting, as it is the marginal cost of production for the government. To see this, consider the government’s cost minimization problem

\[
c(e, p) = \min_{L,T} \quad p_L L + p_T T
\quad \text{s.t.} \quad (\min\{L, T\})^2 \geq e
\]

So, \(L^* = T^* = e^{1/2}\) and we get the value function

\[
C(e) = (p_L + p_T)e^{1/2}
\]
Now, we can calculate the marginal cost

\[
\frac{\partial C}{\partial e} = \frac{1}{2}(p_L + p_T)e^{-1/2}
\]

(8.69)

So, at optimum, \( p_L = p_T = 1 \) and \( e = \frac{4}{9} \), so

\[
MC_e(e^*) = \frac{3}{2}
\]

(8.70)

This is often the solution in trying to decentralize a Pareto optimum. The social planner must compare the cost of production with the benefit that such production gives to the consumer. In the decentralized case, the producer does the job of choosing production and the consumer must be made to clear the market. To reconcile the market to the social planner, we must make both have the same incentives in mind – those of society. Prices are what link the two problems, so it makes sense that the producer’s true marginal costs must be equal to the marginal costs that the agent experiences when deciding what to buy. The way to do this is to set the price of the output equal to its marginal cost and the price of the inputs equal to their marginal productivities.

So, to summarize, under

\[
\begin{align}
  p_T &= p_L = 1 \\
  p_f &= 2 \\
  p_e &= \frac{3}{2} \\
  \tau &= \frac{2}{3}
\end{align}
\]

(8.71) \quad (8.72) \quad (8.73) \quad (8.74)

markets clear when all agents maximize (aside from the government), and we are left with \( e^* = \frac{4}{9} \) and \( f^* = \frac{1}{3} \). So, we have decentralized the Pareto optimum. Government takeover of education has allowed us to reach the social optimum. Of course, how did the government know to set \( e^* = \frac{4}{9} ? \) Something to think about when doing problems like these.

(d) We will continue to assume a “marginal cost price equilibrium”, meaning that the government sells education at its competitive rate (the marginal cost) and funds any deficit it may incur through a non-distortionary (lump-sum) tax. This is a very simple form of \textbf{price-cap regulation} – where the regulator keeps the firm with market power from charging more than marginal cost to consumers and prevents the firm from exiting by subsidizing it. \textbf{Rate-of-return regulation} is an alternative approach where the regulator controls the return on capital gained through profit. These are the two mechanisms most commonly used in regulation of firms with increasing returns to scale (such as utilities).

So, \( p_L = p_T = 1 \) and \( p_f = 2 \) are required from government cost minimization and food industry profit maximization. Say that the government seeks to produce education level \( e \). Since \( e = \min\{L,T\} \), the cost-minimizing government
must choose \( L^* = T^* = e^{1/2} \). So,

\[
C(e) = 2e^{1/2} \tag{8.75}
\]

\[
MC_e(e) = \frac{\partial C}{\partial e} = e^{-1/2} = p_e(e) \tag{8.76}
\]

So,

\[
\Pi_e = p_e e - p_T e^{1/2} - p_L e^{1/2} = e^{-1/2} \cdot e - 2 e^{1/2} = -e^{1/2} \tag{8.77}
\]

Thus, \( \tau = e^{1/2} \) is what the landowner must pay.

(e) At this point, we have already set prices based on firm behavior. Hence, we need to look at consumer behavior. The labor owner solves

\[
\max_{e_l,f_l} e_l^{1/2} f_l^{1/2} \quad \text{s.t.} \quad p_e e_l + p_p f_l \leq 1 \tag{8.78}
\]

The constraint must bind, so \( f_l = \frac{1}{p_f} (1 - p_e e_l) \). Why not just plug in for \( p_e \) now? The agents here are price-takers. Though their behavior affects the price of education, they do not see this and take \( p_e \) as given. Hence, we cannot sub in for \( p_e \) until after the first-order conditions are derived.

\[
\max_{e_l} e_l^{1/2} (1 - p_e e_l)^{1/2} \tag{8.79}
\]

\[
[e_l] : \frac{1}{2} e_l^{-1/2} (1 - p_e e_l)^{1/2} = p_e \frac{1}{2} e_l^{1/2} (1 - p_e e_l)^{-1/2} \tag{8.80}
\]

\[1 - p_e e_l = p_e e_l \tag{8.81} \]

\[e_l^* = \frac{1}{2p_e} \tag{8.82} \]

\[f_l^* = \frac{1}{4} \tag{8.83} \]

But, \( p_e = e^{-1/2} \), so for the labor owner,

\[
(e_l^*, f_l^*) = \left( \frac{1}{2} e^{1/2}, \frac{1}{4} \right) \]

The landowner is similar, but he must pay the tax \( \tau = e^{1/2} \). So,

\[
\max_{e_t,f_t} e_t^{1/2} f_t^{1/2} \quad \text{s.t.} \quad p_e e_t + p_p f_t \leq 1 - \tau \tag{8.84}
\]

Again, \( \tau \) is a given, so we don’t sub in until after the first-order conditions are taken.

\[f_t = \frac{(1 - \tau - p_e e_t)}{p_f} \tag{8.85} \]

\[\text{Notational note: } e \text{ is the government’s planned education production, } e_l^* \text{ and } e_t^* \text{ are the education demands for the landowner and the labor owner, respectively.}\]
So,
\[ \max_{\epsilon_t} \epsilon_t^{1/2} (1 - \tau - p_e \epsilon_t)^{1/2} \]  
(8.86)

\[ [\epsilon_t]: \frac{1}{2} \epsilon_t^{-1/2} (1 - \tau - p_e \epsilon_t)^{1/2} = \frac{1}{2} p_e \epsilon_t^{1/2} (1 - \tau - p_e \epsilon_t)^{-1/2} \]  
(8.87)

1 - \tau - p_e \epsilon_t = p_e \epsilon_t  
(8.88)

\[ \epsilon_t^* = \frac{1 - \tau}{2p_e} \quad f_t^* = \frac{1}{4} (1 - \tau) \]  
(8.89)

Subbing in,
\[ \epsilon_t^* = \frac{1 - \epsilon^{1/2}}{2 \epsilon^{-1/2}} \quad f_t^* = \frac{1}{4} (1 - \epsilon^{1/2}) \]  
(8.90)

\[ (\epsilon_t^*, f_t^*) = \left( \frac{\epsilon^{1/2} - \epsilon}{2}, \frac{1 - \epsilon^{1/2}}{4} \right) \]

So now, we need to government to set \( \epsilon \) to clear markets. We know that inputs used for education are \((L_e, T_e) = (\epsilon^{1/2}, \epsilon^{1/2})\). Hence, inputs for food production are \((L_t, T_t) = (1 - \epsilon^{1/2}, 1 - \epsilon^{1/2})\). So, total education production is \( \epsilon \) and total food production is \( 1 - \epsilon^{-1/2} \). Do markets clear?

\[ \epsilon_t^* + \epsilon_t^* = e = \frac{1}{2} \epsilon^{1/2} - \frac{1}{2} \epsilon + \frac{1}{2} \epsilon^{1/2} \]
\[ f_t^* + f_t^* = \frac{1}{4} - \frac{1}{4} \epsilon^{1/2} + \frac{3}{4} = 1 - \epsilon^{1/2} \]
\[ e^{1/2} = \frac{2}{3} \Rightarrow e = \frac{4}{9} \]

\( e = \frac{4}{9} \)  
(8.91)

So, if total production of education is \( \frac{4}{9} \), and education is priced at is marginal cost, then we have an equilibrium at \( p_T = p_L = 1, p_f = 2, p_e = \frac{4}{3} \). The associated allocations are
\[ (\epsilon_t^*, f_t^*) = \left( \frac{1}{3}, \frac{1}{4} \right) \]

and
\[ (\epsilon_t^*, f_t^*) = \left( \frac{1}{9}, \frac{1}{12} \right) \]

Is this equilibrium Pareto optimal? We must ensure that an all-knowing, all-powerful social planner couldn’t force a change in production and a trade that would make both agents better off. To do this, we must consider two sets.

The first is the production possibility frontier. This is the set of points in \((e, f)\) space where there is no change in production plan that would produce more...
8.3. MARKET FAILURE DUE TO OVERPRODUCTION

Our aggregate endowment of land and labor is $L = T = 1$. So, given this constraint, the production sector can produce $e = (\min\{L,T\})^2$ and $f = (1 - L)^{1/2}(1 - T)^{1/2}$. From here it is clear that $L \neq T$ cannot be optimal, since if we were to reduce $\max\{L,T\}$ to $\min\{L,T\}$, we would not affect education production and would actually increase food production. With $L = T$, we see that the production possibilities frontier is given parametrically by $e = L^2$ and $f = 1 - L$. Eliminating $L$, we find

$$f = 1 - \sqrt{e} \quad (8.92)$$

The other set that we must be concerned with is the set of points in $(e,f)$ space that make both consumers weakly better off. The frontier of this set is known as the Scitovsky contour. In general, it is hard to solve for, but in our case, it is quite easy. Say a consumer with utility $u(e,f) = \sqrt{ef}$ attains utility level $u^*$ with consumption levels $e^*$ and $f^*$. Now say that he has two “children” with the same utility function. If he gives $(\alpha e^*, \alpha f^*)$ to one child and $((1-\alpha)e^*, (1-\alpha)f^*)$ to the other, then he has effectively given them utility levels $\alpha u^*$ and $(1-\alpha)u^*$, respectively. So, the Scitovsky contour in our case is just the indifference curve that runs through the utility level of the aggregate consumption. In our case, aggregate consumption is given by $(4/9, 1/3)$, yielding utility level $u^* = 2\sqrt{3}/3$. The indifference curve running through this point, i.e. the Scitovsky contour, is then given by

$$f = \left(\frac{2}{3\sqrt{3}}\right)^2 \frac{1}{e} = \frac{4}{27e} \quad (8.93)$$

Now, if the Scitovsky contour is above the production possibilities set everywhere but at the aggregate consumption for our marginal pricing equilibrium, then we are Pareto optimal, because there is no different production possibility that can maintain the same level of utility for both the labor owner and the landowner. To see this consider the function $\Delta$ that we construct by subtracting the Scitovsky contour from the production possibilities set,

$$\Delta(e) = \frac{4}{27e} - 1 + \sqrt{e} \quad (8.94)$$

It’s derivative is given by

$$\Delta'(e) = -\frac{4}{27e^2} + \frac{1}{2\sqrt{e}} \quad (8.95)$$

This derivative is positive above $e = 4/9$ and negative below. So, if $\Delta(4/9) = 0$, then we are Pareto optimal. And indeed

$$\Delta(4/9) = \frac{4}{27 \cdot \frac{4}{9}} - 1 + \sqrt{\frac{4}{9}} = \frac{1}{3} - 1 + \frac{2}{3} = 0 \quad (8.96)$$

Hence this equilibrium is Pareto optimal.
Production possibilities frontier is solid,
Scitovsky contour is dashed
Chapter 9

Final Exam Review

9.1 Optimal Durability

9.1.1 Micro comp, June 2005, #1

Consider a profit-maximizing firm that produces a single product. It can choose both the quality of its product and the quantity produced. The firm faces an inverse demand curve $P(q, s)$, where $P_q < 0$ and $P_s > 0$. Its costs are $C(q, s)$, where $C_q, C_s > 0$. This problem asks you to compare the firm’s choice of quantity and quality to the choices that maximize social welfare, defined as the sum of producer and consumer surplus: $\int_0^q P(x, s)dx - C(q, s)$.

(a) Show that for any fixed quantity $q$, the profit-maximizing quality level will typically differ from the socially efficient quality level. Explain the difference.

(b) Provide a simple condition on $P(q, s)$ under which the socially efficient quality exceeds the profit maximizing quality, again conditional on $q$.

(c) Does this condition also imply that the choice of quality is higher when both quantity and quality are chosen to maximize efficiency than when both are chosen to maximize profits? Explain.

Now assume that $s$ reflects the lifetime of the product, so that $q$ units have an effective lifetime of $qs$. Suppose that consumers care about the effective lifetime of the product, so that $P(q, s)/s = p(q, s)$, where $p' < 0$. Suppose also that there are constant returns to scale on the cost side, so that $C(q, s) - q \cdot c(s)$ with $c', c'' > 0$.

(d) Compare the choice of $s$ when $q$ and $s$ are chosen to maximize efficiency to the choice of $s$ when $q$ and $s$ are chosen to maximize profit.

(e) Explain if and how your answer to (d) would change if the cost function satisfied $C(q, s) - \frac{1}{2}(\alpha + s)q^2$ with $\alpha > 0$. 

117
9.1.2 Solution

(a) Define the profit objective by
\[ \Pi(q, s) = qP(q, s) - C(q, s) \]  
(9.1)
and the welfare objective by
\[ W(w, s) = \int_0^q P(x, s)dx - C(q, s) \]  
(9.2)
Then, the monopolist solves
\[ \max_{s \geq 0} \Pi(q, s) \]  
(9.3)
and the social planner solves
\[ \max_{s \geq 0} W(q, s) \]  
(9.4)
Let \( s_m \) be the arg max for the monopolist and \( s_s \) be the arg max for the social planner. Then, first order conditions give us

\[ \frac{\partial \Pi}{\partial s}(s_m) = q \cdot P_s(q, s_m) - C_s(q, s_m) = 0 \]
\[ \frac{\partial W}{\partial s}(s_s) = \int_0^q P_s(x, s_s)dx - C_s(q, s_s) = 0 \]  
(9.5)
These different first order conditions will clearly yield different optima in the general case. The monopolist sets marginal cost equal to \textit{marginal} revenue, while the social planner sets marginal cost equal to \textit{average} revenue. These two objectives clearly provide different incentives.

(b) So, any time the question asks for a “simple condition”, you should suspect that the condition is supermodularity. To compare quality levels, consider the standard discrete choice Topkis setup, \( \max_{s \geq 0} \psi(q, s, \Omega) \), where

\[ \psi(q, s, \Omega) = \Omega qP(q, s) + (1 - \Omega) \int_0^q P(x, s)dx - C(q, s) \]  
(9.6)
Clearly, when \( \Omega = 0 \), this is the social planner’s problem, and when \( \Omega = 1 \), this is the monopolists problem. So, we need a condition that assures us that \( s^* \) is decreasing in \( \Omega \).

\[ \frac{\partial \psi}{\partial \Omega} = qP(q, s) - \int_0^q P(x, s)dx \]
\[ \frac{\partial^2 \psi}{\partial \Omega \partial s} = qP_s(q, s) - \int_0^q P_s(x, s)dx \]  
(9.7)
Topkis then tells us that if \( \frac{\partial^2 \psi}{\partial \Omega \partial s} < 0 \) (i.e. the objective is submodular), then \( s^* \) is weakly decreasing in \( \Omega \). So, our condition then becomes

\[ \int_0^q P_s(x, s)dx > qP_s(q, s) \]  
(9.8)
or, more intuitively
\[
\frac{1}{q} \int_0^q P_s(x, s) \, dx > P_s(q, s)
\]  
(9.9)

So, what must \( P_s \) be doing on the interval \([0, q]\) for its average value to exceed its endpoint value? It must be decreasing in \( q \). Hence, the condition that ensures that the optimum social quality is weakly greater than the optimum monopoly quality is
\[
P_{sq}(q, s) \leq 0
\]  
(9.10)

(c) Now, we are “unfreezing” quantity, so we need supermodularity in \((\Omega, q, s)\) and not just in \((\Omega, s)\). So, we should start calculating some mixed partials. The \((\Omega, q)\) partial yields
\[
\frac{\partial^2 \psi}{\partial \Omega \partial q} = P(q, s) + q P_q(q, s) - P(q, s) = q P_q(q, s) < 0
\]  
(9.11)

where the inequality comes from our assumption that \( P_q > 0 \) (i.e. the “law of demand”). OK so far. The \((q, s)\) partial yields
\[
\begin{align*}
\frac{\partial \psi}{\partial s} &= \Omega q P_s(q, s) + (1 - \Omega) \int_0^q P_s(x, s) \, dx - C_s(q, s) \\
\frac{\partial^2 \psi}{\partial s^2} &= \Omega P_s(q, s) + \Omega q P_{sq}(q, s) + (1 - \Omega) P_s(q, s) - C_{sq}(q, s) \\
&= \underbrace{P_s(q, s)}_{(+)} + \underbrace{\Omega q P_{sq}(q, s)}_{(-)} - \underbrace{C_{sq}(q, s)}_{(?)} 
\end{align*}
\]  
(9.12)

So, we cannot sign this quantity because we do not know the sign of \( C_{sq} \). Hence, we cannot make the same conclusion that we made in (b).

(d) So, now we enter the part of the problem that is concerned with durability. What is important to consumers is now the amount of time their stuff will last. It makes no difference to them if they buy ten transformers that last a year a piece for $1 a piece or if they buy one transformer for $10 that lasts ten years. Since lifetime \((y \equiv qs)\) is so important now, let’s try to solve the optimization problem in terms of lifetime \( y \) and durability \( s \). So, we go from
\[
\Pi = qsp(qs) - qc(s) \\
W = \int_0^q sp(xs) \, dx - qc(s)
\]  
(9.13)

to
\[
\begin{align*}
\Pi &= yp(y) - y \frac{c(s)}{s} \\
W &= \int_0^y p(\eta) \, d\eta - y \frac{c(s)}{s}
\end{align*}
\]  
(9.14)

by using the transformation \( q = y/s \) \((dq = dy/s)\).

Note that the part of the objective that is concerned with \( s \) is the same for both the monopolist and the social planner, and that it is quite segregated from any mention of \( y \). In fact, for any value of \( y \), both agents will prefer to set \( s \) such that it minimizes the average cost, \( c(s)/s \). So the optimal durability is the same for the monopolist and the social planner. In other words, left to his devices, the monopolist will produce at the socially optimal level of durability.
(e) Before, the per unit cost was a function of the durability of the good. Now, it is more expensive to produce many units, so *ceteris paribus*, a firm prefers making a $10 transformer that lasts 10 years to making 10 $10 transformers that last a year a piece. Let’s write this new cost function in terms of $y$.

\[ C(q, s) = \frac{1}{2}(\alpha + s)q^2 \]

then becomes

\[ C\left(\frac{y}{s}, s\right) = \frac{1}{2}(\alpha + s)\left(\frac{y}{s}\right)^2 = \frac{1}{2}y^2\left(\frac{\alpha}{s^2} + \frac{1}{s}\right) \tag{9.15} \]

Our objectives are then

\[ \Pi = yp(y) - \frac{1}{2}y^2\left(\frac{\alpha}{s^2} + \frac{1}{s}\right) \]

\[ W = \int_0^y p(\eta) d\eta - \frac{1}{2}y^2\left(\frac{\alpha}{s^2} + \frac{1}{s}\right) \tag{9.16} \]

Again, $s$ is segregated, so both the monopolist and the social planner will choose the same durability (in this case $s^* = \infty$). So nothing much has changed, but clearly the theorem proved in (d) begins to have problems when increasing marginal costs kick in. All of the sudden the optimum is to produce $\varepsilon$ transformers that lasts for $\infty$ years, where $\varepsilon \cdot \infty$ equals some finite number $y$. This means that our assumptions of infinite divisibility and consumers only being concerned with the lifetime of the product have come back to bite us. Clearly, if 100 consumers need a combined transformer lifetime of 100 years, their problem is not solved by the production of 1 transformer that lasts 100 years. This is just an illustration of the theorem shown in part (d) (Swan’s Optimum Durability Theorem).

### 9.2 Labor managed firms

#### 9.2.1 Micro comp, June 2005, #1

Consider a world where capital and labor are inputs to production. Given an output price $p$ and input prices $w, r$, the standard *neoclassical firm* chooses inputs to maximize profits:

\[ \max_{k, l \geq 0} pf(k, l) - wl - rk \tag{9.17} \]

The firm will choose to operate if and only if it can achieve non-negative profits.

In contrast, the objective of a *labor-managed firm* is to maximize profits *per unit of labor*:

\[ \max_{k, l \geq 0} \frac{1}{l}[pf(k, l) - wl - rk] \tag{9.18} \]

If a firm is labor-managed, it will operate if and only if it can give a revenue share to each worker that exceeds forgone wages (i.e. if the solution to its maximization problem is non-negative).

Assume that $f(\cdot)$ is continuous, increasing, and differentiable, that $f(0, 0) = 0$, and that solutions to the problems you encounter are unique.

For parts (a)-(c) assume that capital is fixed at some level $K > 0$. 


(a) Show that in response to a higher product market price $p$, the neoclassical firm will raise its labor input.

(b) How will a labor managed firm respond to this same increase, assuming that $f(\cdot)$ is concave.

(c) Show that if prices $(p, w, r)$ are such that a labor-managed firm would operate at positive levels, then the labor managed firm will choose a lower level of labor input than the neoclassical firm.

(d) Are your conclusions to (a)-(c) still valid if capital is a choice variable?

(e) Show that the set of price $(p, w, r)$ at which the neoclassical and labor-managed firms would operate is identical.

(f) Paul Samuelson wrote that “in a perfectly competitive economy, it does not matter whether capital hires labor or labor hires capital.” Show that if prices $(p, w, r)$ are such that the neoclassical firm earns zero profits, the neoclassical and labor-managed firms would make identical choices.

9.2.2 Solution

(a) So, the neoclassical firm solves (in the “short run”)

$$
l^* (k, w, r) \in \arg \max_{l \geq 0} \underbrace{pf(k, l) - wl - rk}_{\equiv \Pi_n (k, l)}
$$

(9.19)

We are concerned about how the arg max changes in the parameters, so Topkis’ theorem is the way to proceed.

$$
\frac{\partial \Pi_n}{\partial p} (k, w, r) = f(k, l)
\frac{\partial^2 \Pi_n}{\partial p \partial l} (k, w, r) = f_l (k, l) > 0
$$

(9.20)

Our objective is supermodular in $(p, l)$, thus $l^*$ is weakly increasing in $p$.

(b) The labor-owned firm solves

$$
l^* (k, w, r) \in \arg \max_{l \geq 0} \underbrace{pf(k, l) - wl - rk}_{\equiv \Pi_l (k, w, r)}
$$

(9.21)

Again, we will use Topkis. Calculating the mixed partial

$$
\frac{\partial \Pi_l}{\partial p} (k, w, r) = \frac{f(k, l)}{l}
\frac{\partial^2 \Pi_l}{\partial p \partial l} (k, w, r) = \frac{1}{l} \left[ f_l (k, l) - \frac{f_l (k, l)}{l} \right]
$$

(9.22)

How do we sign this? The two terms in the square brackets represent the marginal product of labor and the average product of labor. Since $f_l$ is concave, we know that $f_l$ is decreasing. So, $f_l$ at an endpoint will be less than the average of $f_l$ over an interval. Graphically,
Clearly \( \frac{f'}{l} > f_l \). So, \( \frac{\partial^2 \Pi}{\partial p \partial l} < 0 \).

Actually, we can do this a bit more rigorously with the mean value theorem. If we have a differentiable function \( f \) on the interval \([a, b]\), then \( \exists c \in [a, b] \) s.t.

\[
f'(c) = \frac{f(b) - f(a)}{b - a}
\]

(9.23)

If we manipulate this, we get something vaguely reminiscent of Taylor’s theorem

\[
f(b) = f(a) + f'(c)(b - a)
\]

(9.24)

We call this a **mean value expansion**. Applying it to \( f \) on \([0, l]\), we find

\[
f(\bar{k}, l) = f(\bar{k}, 0) + f'(l^\circ)l \quad \quad (l^\circ \in [0, l])
\]

(9.25)

Plugging this mean value expansion in to the expression we calculated for the mixed partial yields

\[
\frac{\partial^2 \Pi}{\partial p \partial l} = \frac{1}{l} \left[ f_l(\bar{k}, l) - \frac{f(\bar{k}, l)}{l} \right] = \frac{1}{l} \left[ f_l(\bar{k}, l) - f_l(\bar{k}, l^\circ) \right]
\]

(9.26)

\( f \)'s concavity tells us that \( f_l \) is decreasing in \( l \), and since \( l^\circ \leq l \), we know that \( l^\circ \leq l \). Hence, \( f_l(\bar{k}, l) - f_l(\bar{k}, l^\circ) < 0 \) and we conclude that \( \frac{\partial^2 \Pi}{\partial p \partial l} < 0 \). Topkis then tells us that \( l^* \) is weakly decreasing in \( p \).

(c) Note that applying a log transformation to the objective will not affect the arg max. So, we can write both firm’s problems as

\[
l^*(\bar{k}, w, r, \Omega) \in \arg \max_{l \geq 0} \left[ pf(\bar{k}, l) - wl - rk \right] = \Omega \log l
\]

(9.27)

where \( \Omega = 0 \) represents the neoclassical problem and \( \Omega = 1 \) represents the labor-managed problem. Now, we calculate the mixed partial so that we may apply Topkis.

\[
\frac{\partial \psi}{\partial l} = -\log l, \quad \frac{\partial^2 \psi}{\partial l^2} = -\frac{1}{l} < 0
\]

(9.28)
So, the objective is submodular in \((l, \Omega)\), meaning that \(l^*\) is weakly decreasing in \(\Omega\).

Hence, the labor-managed firm hires less labor than the neo-classical firm.

(d)(d-a) Now, we need supermodularity in \((p, k, l)\). We have already checked the mixed partial \((p, l)\), so now we must check the mixed partials \((p, k)\) and \((k, l)\). Starting with \((p, k)\), we find

\[
\frac{\partial^2 \Pi_n}{\partial p \partial k} = f_k(k, l) > 0
\]

where the inequality holds because the production function is increasing in both its arguments. Calculating \((k, l)\),

\[
\frac{\partial \Pi_n}{\partial k} = pf_l(k, l) - w \quad \frac{\partial^2 \Pi_n}{\partial k \partial l} = pf_{kl}(k, l)
\]

So, supermodularity in \((p, k, l)\) requires that \(f_{kl} > 0\). If this is not true, then our conclusion in (a) is no longer true.

(d-b) Again, we have to check the \((p, k)\) and \((k, l)\) mixed partials. Starting with \((p, k)\), we find

\[
\frac{\partial^2 \Pi_l}{\partial p \partial k} = \frac{f_k(k, l)}{l} > 0
\]

where the inequality follows from the fact that the production function is increasing in both arguments. Now, we check the \((l, k)\) mixed partial

\[
\frac{\partial \Pi_l}{\partial k} = pf_k(k, l) - w \quad \frac{\partial^2 \Pi_l}{\partial k \partial l} = pf_{kl}(k, l) - pf_k(k, l) - \frac{r}{l}
\]

Again, we can’t sign this, so the conclusion from (b) does not hold. If we had \(f_{kl} > \frac{r}{l}\), we could maintain our conclusion from (b), though this would imply making an assumption about \(f_{kl}\) (through a simple argument using mean value expansions).

(d-c) So, we must calculate more mixed partials of the \(\psi\) function that we used in (c). Calculating the \((\Omega, k)\) mixed partial, we find

\[
\frac{\partial^2 \psi}{\partial \Omega \partial k} = 0
\]

So we are fine there. Calculating the \((k, l)\) mixed partial, we find

\[
\frac{\partial \psi}{\partial k} = \frac{pf_k(k, l) - r}{pf(k, l) - wl - rk} \quad \frac{\partial^2 \psi}{\partial k \partial l} = \frac{pf_{kl}(k, l)}{pf(k, l) - wl - rk} - \frac{pf_k(k, l) - r}{(pf(k, l) - wl - rk)^2} \left( pf_l(k, l) - w \right)
\]

Again, the sign of this mixed partial is unclear given our assumptions, so we cannot make the same conclusion that we made in (c).
(e) If $\Pi_n > 0$, then so is $\Pi_l$. The same holds in reverse. So, if positive profit is possible for one firm, it is possible for both. Hence both firms will choose to operate at the same set of $(p, w, r)$ parameters.

(f) Say $(k^*, l^*) = \arg \max_{k,l \geq 0} \Pi_n$. $(k^*, l^*)$ yield zero profits, so changing them at all leads to $\Pi_n < 0$. But, $\Pi_l = \frac{\Pi_n}{\ell}$, so if $\Pi_n < 0$, so is $\Pi_l$. Hence, if the neoclassical firm earns zero profits at optimum, then the labor-owned firm also earns zero profits at optimum, and as such, must choose the same $(k^*, l^*)$.

### 9.3 Continuous GE

#### 9.3.1 2005 Econ 202 Final, #3

Consider an economy with two goods: food and club memberships. Each agent has identical preferences $u(x, y) = (1 + \lambda y)x$, where $x \geq 0$ is food consumption, $y \in \{0, 1\}$ is club membership, and $\lambda > 0$. There is a unit mass of agents, of which half are endowed with club memberships, so the total mass of memberships is $1/2$. The agents endowed with memberships and those not endowed with memberships both have food endowments uniformly distributed on $[0, 1]$. Suppose the agents can trade both food and club memberships. Normalize the price of food $p_x = 1$ and let $p_y = p$ denote the price of club memberships. (Note that with a continuum of agents, market clearing means that the mass of buyers is just equal to the mass of sellers.)

(a) Derive each agent’s Marshallian demand as a function of his endowment.

(b) Explain why in a Walrasian equilibrium there must be trade in club memberships.

(c) Compute the Walrasian equilibrium prices and allocation.

(d) Suppose instead that $u(x, y) = (1 + \lambda y^2)x$, where $Y$ is the aggregate consumption of club memberships, so that club memberships become less valuable as more people consume them. Prove that the Walrasian equilibrium allocation in this setting is not Pareto efficient, i.e. that the first welfare theorem fails.

#### 9.3.2 Solution

(a) Consider an agent with wealth $w$. His budget constraint will bind (the objective is locally nonsatiated), so $x(y) = w - py$. So, now how do we make the decision with $y$ being a discrete choice? We just find the utility in both cases and compare. So

- $u_0 = u(x(0), 0) = x(0) = w$
- $u_1 = u(x(1), 1) = (1 + \lambda)x(1) = (1 + \lambda)(w - p)$
So, the agent will choose to attain a club membership if \( u_1 \geq u_0 \), or

\[
\begin{align*}
  w - p + \lambda w - \lambda p & \geq w \\
  \lambda w & \geq (1 + \lambda)p
  \end{align*}
\]

(9.35)

Now, if \( \omega \) is the food endowment and \( \eta \) is the membership endowment, then \( w = \omega + p\eta \). Hence, we can write Marshallian demand as

\[
(x, y)(\omega, \eta) = \begin{cases} 
  (\omega + p, 0) & \eta = 1, \omega \in [0, \frac{p}{\lambda}] \\
  (\omega, 1) & \eta = 1, \omega \in \left[\frac{p}{\lambda}, 1\right] \\
  (\omega, 0) & \eta = 0, \omega \in \left[0, \frac{1 + \lambda}{\lambda} p\right] \\
  (\omega - p, 1) & \eta = 0, \omega \in \left[\frac{1 + \lambda}{\lambda} p, 1\right]
\end{cases}
\]

(9.36)

(b) So, for there to be no trade in memberships, we clearly require that \( y(\omega, 1) = 1, \forall \omega \). For this to happen, we need \( \frac{p}{\lambda} \leq 0 \), or \( p = 0 \). But, if this is the case, then everyone will demand a membership. Since the measure of memberships is 1/2 and the measure of agents is 1, this cannot be in equilibrium.

(c) The total measure of agents who demand a membership under price \( p \) is

\[
M_D = \frac{1}{2} \left(1 - \frac{p}{\lambda}\right) + \frac{1}{2} \left(1 - \frac{1 + \lambda}{\lambda} p\right) = 1 - \frac{1}{2} \left(\frac{2 + \lambda}{\lambda} p\right)
\]

(9.37)

The total measure of the supply of memberships is

\[
M_S = \frac{1}{2}
\]

(9.38)

Setting \( M_D = M_S \) clears the market, yielding the equilibrium price

\[
\hat{p} = \frac{\lambda}{\lambda + 2}
\]

(9.39)

Plugging this into the Marshallian demand yields the Walrasian allocation

\[
(x, y)(\omega, \eta) = \begin{cases} 
  (\omega + \frac{\lambda}{\lambda + 2}, 0) & \eta = 1, \omega \in [0, \frac{1}{\lambda + 2}] \\
  (\omega, 1) & \eta = 1, \omega \in \left[\frac{1}{\lambda + 2}, 1\right] \\
  (\omega, 0) & \eta = 0, \omega \in \left[0, \frac{1 + \lambda}{\lambda + 2}\right] \\
  (\omega - \frac{\lambda}{\lambda + 2}, 1) & \eta = 0, \omega \in \left[\frac{1 + \lambda}{2 + \lambda}, 1\right]
\end{cases}
\]

(9.40)

We see that the food market clears by seeing that the top and bottom cases in the boxed equation above trade with each other while the middle cases do not trade. The measure of the agents in the first case is \( \frac{1}{\lambda + 2} \). The measure of the agents in the last (fourth) case is the same \( 1 - \frac{1 + \lambda}{2 + \lambda} = \frac{1}{\lambda + 2} \).
(d) Clearly the externality could cause the welfare theorem to fail. We prove this by constructing an explicit Pareto improvement over the Walrasian equilibrium under externality (at least conditionally).

Fortunately, the Walrasian equilibrium is the same that we solved in part (c), since the fact that everyone is maximizing assuming that \( Y = 1/2 \) exogenously (they see \( Y \) as an externality) leads to \( Y \) being 1/2 endogenously. \( Y = 1/2 \) is a self-fulfilling prophecy, and thus the Walrasian equilibrium is the same as before.

Now, how can we show that the first-welfare theorem fails? The easiest way is to explicitly construct a Pareto improvement. The problem in this market is that membership owners hurt other membership owners by coming to the club. The simple solution is for the members that get the most return from the club to pay those who get the least to stay out. Return to club membership is \( \lambda \) times consumption, so those who have memberships in equilibrium with the lowest corresponding consumptions are the ones we would like to pay to stay out. Say that we pay off a small chunk of people who were endowed with a membership, so that now only those with \( \eta = 1, \omega \in \left[ \frac{1}{\lambda + 1}, \frac{1}{\lambda + 1} + 2\varepsilon \right] \) or \( \eta = 0, \omega \in \left[ \frac{\lambda + 1}{\lambda + 2}, 1 \right] \) actually come to the club. Hence, the new total measure of members who come to the club is

\[
Y = \frac{1}{2} \left( 1 - \frac{1}{\lambda + 2} - 2\varepsilon + 1 - \frac{\lambda + 1}{\lambda + 2} \right) = \frac{1}{2} (1 - 2\varepsilon)
\]

(9.41)

Those in the chunk that are payed \( \pi \) to stay out of the club must be better off to accept such an arrangement. Hence, to maintain Pareto neutrality we need

\[
(1 + \lambda) \frac{1}{\lambda + 2} \leq \frac{1}{\lambda + 2} + \pi
\]

or

\[
\pi \geq \frac{\lambda}{\lambda + 2}
\]

(9.42)

(9.43)

It makes sense that the payment must be at least the price under Walrasian equilibrium, since the price is designed to make the marginal consumer indifferent between buying and not buying a club membership.

The measure of the chunk we are paying off is \( \varepsilon \), so it will cost us \( C = \frac{\varepsilon}{\lambda + 2} \) in food to run our scheme. Where will we get the funds?

Let’s determine how much we could possibly charge the club members so that they still prefer paying the fee to having the \( \varepsilon \) chunk using their club? For a fee \( \kappa \) and a consumption \( c \), this requires

\[
(1 + \lambda)c \leq \left( 1 + \frac{\lambda}{2(\frac{1}{2} - \varepsilon)} \right) (c - \kappa)
\]

(9.44)

To make things more simple, let’s use the fact that \( \varepsilon \) is infinitesimal. Taylor expanding to first-order\(^1\), we find

\[
(1 + \lambda)c \leq (1 + \lambda + 2\lambda\varepsilon)(c - \kappa)
\]

(9.45)

\(^1\)\((1 + \varepsilon)^n \approx 1 + n\varepsilon \) for small \( \varepsilon \) is a good approximation to know off-hand. So is \( e^\varepsilon \approx 1 + \varepsilon \).
Solving for $\kappa$ yields

$$\kappa \leq \frac{2\varepsilon \lambda}{1 + \lambda + 2\varepsilon \lambda} c$$  \hspace{1cm} (9.46)

Again, we Taylor expand in order to keep everything linear in $\varepsilon$, yielding

$$\kappa \leq \frac{2\varepsilon \lambda}{1 + \lambda} c$$  \hspace{1cm} (9.47)

Now, we can integrate to determine how much food we could take from the remaining club members in order to pay off the $\varepsilon$ chunk that we are excluding.

$$R = \frac{1}{2} \left( \int_{\frac{1}{\lambda + 2}}^{1} \frac{2\varepsilon \lambda}{1 + \lambda} \omega d\omega + \int_{\frac{\lambda + 1}{\lambda + 2}}^{1} \frac{2\varepsilon \lambda}{1 + \lambda} \left( \omega - \frac{\lambda}{\lambda + 2} \right) d\omega \right)$$  \hspace{1cm} (9.48)

After some grunge, this comes out to

$$R = \frac{\varepsilon}{2} \left( \frac{3\lambda}{(1 + \lambda)(2 + \lambda)^2} + \frac{\lambda(3 + \lambda)}{(2 + \lambda)^2} \right) = \frac{\varepsilon}{2} \left( \frac{3\lambda + \lambda(3 + \lambda)(1 + \lambda)}{(1 + \lambda)(2 + \lambda)^2} \right)$$  \hspace{1cm} (9.49)

So, if $R > C$, then we have found our Pareto improvement – we have made everyone just as well off and have food leftover to make people even better off. Is it?

$$R > C$$

$$\frac{\varepsilon}{2} \left( \frac{3\lambda + \lambda(3 + \lambda)(1 + \lambda)}{(1 + \lambda)(2 + \lambda)^2} \right) > \frac{\kappa c}{\lambda + 2}$$

$$\frac{1}{2} \left( \frac{3 + (3 + \lambda)(1 + \lambda)}{(1 + \lambda)(2 + \lambda)} \right) > 1$$  \hspace{1cm} (9.50)

$$3 + (3 + \lambda)(1 + \lambda) > 2(1 + \lambda)(2 + \lambda)$$

$$6 + 4\lambda + \lambda^2 > 4 + 6\lambda + 2\lambda^2$$

$$2 > 2\lambda + \lambda^2$$

So the revenue we collect from club members is enough to payoff the marginal membership from coming to the club only if $\lambda < \sqrt{3} - 1 \approx 0.732$. So, in spite of what the question tells you to prove, you can only come up with a Pareto improvement if $\lambda$ is sufficiently small. The general proof is too complicated to present in a section, but the moral here is that the questions aren’t always solvable without a little change. In this case, the question needs, “for sufficiently small $\lambda$”.

This answer is quite sensible. As a social planner, as $\lambda$ gets larger, the hit you take from not having people use the club (and get an extra $\lambda$ return on their food) is too large. Hence, the number of socially optimal club-goers should increase with $\lambda$. At some point then, it makes sense that $\lambda$ would be so high that $Y > 1/2$ would be optimal if it were feasible. The social planner then hits a constraint and $Y = 1/2$ becomes the optimum. If the externality is held at the socially optimal level, then it no longer poses a problem for the welfare theorem, as we have just shown.
9.4 Monopolists, Social Planners, and Taxes

9.4.1 Question #2, 202 Final 2005

A monopolistic firm sells its output $y$ to $N$ consumers, whose utilities are functions of their consumptions of good $y$ and the numeraire good, and are quasilinear in the numeraire good. Consumers make their purchase decisions taking the price $p$ set by the firm as given (All prices will be in terms of the numeraire good.) Let $Q(p)$ denote the consumers’ aggregate demand for good $y$ as a function of its price $p$.

The firm’s production function is $f(k, l)$, where $k$ and $l$ are the inputs of capital and labor, respectively, purchased at prices $r$ and $w$, respectively. Assume that the production function is increasing and has constant returns to scale.

The firm chooses price $p$ to maximize its profits in terms of the numeraire good.

(a) Derive the total Marshallian surplus of the consumers as a function of the price $p$ set by the firm.

(b) Compare the output resulting from the firm’s profit maximization pricing to the Pareto efficient output level.

Suppose now that the government levies a tax $\tau$ on each unit of output bought by the consumers, so the total price consumers pay becomes $p + \tau$. You are asked to make monotone comparative statics predictions about the effects of the tax. In some of the questions, you may need to make additional assumptions on the production function or the demand curve to make such a prediction. How does the tax rate $\tau$ affect

(c) the firm’s output,

(d) its demands for capital and labor,

(e) the price it sets?

Now suppose that we observe the firm’s profit-maximizing output $y(\tau)$ at all tax rates $\tau$.

(f) Calculate the firm’s cost function.

(g) Is observation of $y(\tau)$ enough to fully infer the firm’s production technology?

9.4.2 Solution

(a) Say we are considering a production level $q$. The revenue that the firm makes from selling this amount is $R = qp$ where $p$ is the price associated with $q$ by the demand curve. But what if the firm were able to run the following scheme. First it would set $p = \infty$ and sell to whoever is willing to pay such a price. Then, it would sell at $p = \infty - dp$, and so on. Essentially, this is a scheme that allows the firm to have perfect price discrimination.
How much could the firm make in this case? Well, at $p = \infty$, the firm sells to zero consumers. Then, the firm lowers its price to $P(dQ)$ so that it may sell to $dQ$ consumers, making $P(dQ)dQ$ in revenue. Then, it lowers its price some more, to $P(2dQ)$ and is able to sell to $2dQ$ consumers. But it already sold to $dQ$ of them so it only sells to the new $dQ$, making $P(2dQ)dQ$ in revenue. Continuing in this manner, it is clear that the revenue the firm can make in discriminating by selling $q$ total units is $R_d = \int_0^q P(\tilde{q})d\tilde{q}$.

The difference between these revenues, $R_d - R$, represents gains that the consumers make from trade. This is the consumer surplus, $CS = R_d - R$. Unfortunately, it is given in terms of $q$ and not $p$. Is it possible to express it in terms of $p$. Consider integration by parts

\begin{align*}
CS &= R_d - R \\
CS &= \int_0^q P(\tilde{q})d\tilde{q} - qP(q) \\
CS &= \tilde{q}P(\tilde{q})|_0^q - \int_0^q \tilde{q}P'(\tilde{q})d\tilde{q} - qP(q) \\
CS &= qP(q) + \left. \frac{1}{2}P(0) \right|_0^q - \int_0^q \tilde{q}P'(\tilde{q})d\tilde{q} - qP(q) \\
CS &= -\int_0^q \tilde{q}P'(\tilde{q})d\tilde{q}
\end{align*}

followed with the substitution $\tilde{p} = P(\tilde{q}) \ (d\tilde{p} = P'(\tilde{q})d\tilde{q} ; \tilde{q} = Q(\tilde{p}))$. Then, we find

\begin{align*}
CS &= -\int_{P(0)}^{P(q)} Q(\tilde{p})d\tilde{p} = -\int_{\infty}^P Q(\tilde{p})d\tilde{p} = \int_{P}^{\infty} Q(\tilde{p})d\tilde{p} \quad (9.52)
\end{align*}

So, the Marshallian surplus is best expressed as

\[ CS = \int_{P}^{\infty} Q(\tilde{p})d\tilde{p} \]

Note that this could have been equally well seen graphically. The consumer surplus is the area between the demand curve and the price line, as shown below. Clearly this area can be equally well expressed as the area to the left of the demand curve and above the price line. Hence $\int_{P}^{\infty} Q(\tilde{p})d\tilde{p} = \int_0^q P(\tilde{q})d\tilde{q} - qP(q)$. 

![Diagram showing consumer surplus (CS) as the area between the demand curve $P(q)$ and the price line $p$.](image)
(b) Define the cost function of the firm by
\[
C(q, w, r) = \min_{k,l \geq 0} \ w l + rk \\
\text{s.t. } f(k, l) \geq q
\]
(9.53)

The firm’s profit maximization problem is given by
\[
\Pi(w, r) = \max_{p \geq 0} \ p Q(p) - C(Q(p), w, r)
\]
(9.54)

Now, the social planner maximizes the sum of firm profits and consumer surplus. Hence, the firm’s and social planner’s problems can be united via the parameter \(\Omega\) in the optimization
\[
p^*(w, r, \Omega) = \arg \max_{p \geq 0} \ p Q(p) - C(Q(p), w, r) + \Omega \int_p^\infty Q(\tilde{p})d\tilde{p}
\]
≡ \(\varphi\)
(9.55)

Now, we can determine whether the social planner or the monopolist sets a higher price. Calculating cross-partial
\[
\frac{\partial \varphi}{\partial \Omega} = \int_p^\infty Q(\tilde{p})d\tilde{p}
\]
\[
\frac{\partial^2 \varphi}{\partial \Omega \partial p} = -Q(p) < 0
\]
(9.56)

So, since the cross-partial is negative, Topkis tells us that \(p^*\) is decreasing in \(\Omega\). \(\Omega = 0\) represents the monopolist’s problem, while \(\Omega = 1\) represents the social planner’s. Hence \(p^*(0) \geq p^*(1)\) tells us that \(p^*_M \geq p^*_SP\). And finally, since \(Q(p)\) is decreasing\(^2\), we make the conclusion that the monopolist underproduces relative the social optimum,
\[
q^*_M \leq q^*_SP
\]
(9.57)
as we all remember from introductory microeconomics.

(c) First, note that a constant-returns-to-scale, single-output production function guarantees us a cost-function that is homogeneous of degree one. To see this, consider
\[
C(\lambda q) = \arg \min_{k,l \geq 0} \ w l + rk = \lambda \arg \min_{\frac{k}{\lambda}, \frac{l}{\lambda} \geq 0} \ w \frac{k}{\lambda} + r \frac{l}{\lambda} = \lambda C(q) \\
\text{s.t. } f(k, l) \geq \lambda q \\
\text{s.t. } f\left(\frac{k}{\lambda}, \frac{l}{\lambda}\right) \geq q
\]
(9.58)

And, a univariate function that is homogeneous of degree one must be linear. To see this, consider the converse of Euler’s theorem. It provides a differential equation for \(C(q)\).
\[
qC'(q) = C(q)
\]
(9.59)

\(^2\)Since the consumer has quasilinear preferences, there can be no wealth effects, and the Marshallian demand is the same as the Hicksian demand. Hence, the Marshallian demand must obey the law of demand, i.e. \(Q' < 0\).
The general solution to this equation is given by
\[
\frac{dC}{C} = \frac{dq}{q} \\
\log C = \log q + \log K \\
C(q) = Kq
\] (9.60)

where \( K \) is a constant of integration, subject to the initial conditions of the problem.

With this piece of knowledge, we can now write the firm’s profit maximization problem as
\[
\Pi(w, r, \tau) = \max_{p \geq 0} \underbrace{pQ(p + \tau) - K(w, r)Q(p + \tau)}_{\equiv \varphi}
\] (9.61)

Calculating cross-partial,
\[
\frac{\partial \varphi}{\partial \tau} = pQ'(p + \tau) - K(w, r)Q'(p + \tau) \\
\frac{\partial^2 \varphi}{\partial \tau \partial p} = Q'(p + \tau) + pQ''(p + \tau) - K(w, r)Q''(p + \tau) \\
= Q'(p + \tau) + (p - K(w, r))Q''(p + \tau)
\] (9.62)

How can we sign this mixed partial. The \( Q' \) term is negative. The firm’s profit objective can be written as \( Q(p - K) \), so if the firm is choosing not to shut down, then it must be that \( p > K \). Hence, if we assume that \( Q'' < 0 \), then we can sign the mixed-partial to be negative. Topkis would then tell us that \( p^*(w, r, \tau) \) is decreasing in \( \tau \). And since \( Q(p) \) is decreasing, we must find that \( Q(p^*(w, r, \tau)) \) is increasing with \( \tau \). So, granted that \( Q'' < 0 \),

The firm’s output increases with the imposition of the tax.

(d) If \( Q'' < 0 \), then output will increase with the imposition of the tax. And since the production function is constant-returns-to-scale, the arg maxes from producing \( \lambda q \) are just \( \lambda \) times the arg maxes from producing \( q \). So, if production increases with the imposition of the tax, so will factor demands. Hence, granted that \( Q'' < 0 \),

The demands for capital and labor will go up with the imposition of the tax.

(e) As we saw in part (c), assuming that \( Q'' < 0 \),

The price decreases with the imposition of the tax.

(f) If output \( y \) is produced, then the market-clearing price is given by the inverse demand \( P(y) \) (\( \equiv Q^{-1} \)). But \( \tau \) of this is tax. Phrasing the firm’s optimization in terms of output \( y \), we find
\[
\Pi(\tau) = \arg \max_y [P(y) - \tau - K]y
\] (9.63)

Hence, by the envelope theorem,
\[
\Pi'(\tau) = -y(\tau)
\] (9.64)
or by the integral envelope theorem
\[ \Pi(\infty) - \Pi(\tau) = -\int_{\tau}^{\infty} y(\tilde{\tau}) d\tilde{\tau} \]  
(9.65)

where we cancelled \( \Pi(\infty) \) because no one will purchase the good if it has an infinite tax on it. So, we are left with
\[ \Pi(\tau) = \int_{\tau}^{\infty} y(\tilde{\tau}) d\tilde{\tau} \]  
(9.66)

Plugging the arg max into the objective, we also have
\[ \Pi(\tau) = [P(y(\tau)) - \tau - K]y(\tau) \]  
(9.67)

Solving for \( K \) and plugging in the envelope identity, we attain an expression for \( K \),
\[ K = P(y(\tau)) - \tau - \frac{\Pi(\tau)}{y(\tau)} = P(y(\tau)) - \tau - \frac{1}{y(\tau)} \int_{\tau}^{\infty} y(\tilde{\tau}) d\tilde{\tau} \]  
(9.68)

Hence, cost is given by
\[ C(q) = Kq = \left( P(y(\tau)) - \tau - \frac{1}{y(\tau)} \int_{\tau}^{\infty} y(\tilde{\tau}) d\tilde{\tau} \right) q \]  
(9.69)

(g) Since, we have the cost function and the solution to the profit maximization problem, so, via Shepard’s lemma, we can express the unconditional factor demand as
\[ Y^*(r, w, \tau) = \left( 1, \frac{\partial K}{\partial r}(r, w), \frac{\partial K}{\partial l}(r, w) \right) y(\tau) \]  
(9.70)

Beyond this, it is hard to proceed without extra information. Remember the solution to the producer theory set, problem number 4. All three parts had production functions of form \( C(w, r) = K(w, r)q \). On some we could completely determine the production set, and on some we could not. Hence, the answer to this problem depends on what kind of \( y(t) \) we are given, and what the resultant \( K(w, r) \) that we calculate from it, via (f).
Appendix A

Duality theorem

A.1 The duality theorem

A buzz word we hear quite regularly in graduate microeconomics is dual. So, what exactly is the dual problem? In terms of what we are learning right now (producer and consumer theory), it seems to refer to attacking the same problem from two different perspectives. For instance, consider the utility maximization problem (UMP):

\[
v(p, w) = \max_x u(x) \quad \text{s.t.} \quad p \cdot x \leq w
\]  

(A.1)

What is the other way that we approach the consumer’s problem (CP)? With the expenditure minimization problem (EMP):

\[
e(p, u) = \min_x p \cdot x \quad \text{s.t.} \quad u(x) \geq u
\]  

(A.2)

Both optimization problems capture what the consumer must do given either a) a target utility, or b) a wealth level. The similarities run even deeper than this.

The UMP has maximization of a non-linear utility objective and a linear price constraint. The EMP is essentially the same problem turned on its head - minimization of a linear price objective and a non-linear utility constraint. This is at the heart of the idea of the meaning of dual - to recast the problem with the objective as a constraint and the constraint as the objective.

Another property of these problems with far-reaching consequences is the linearity of pricing in them. This linearity, either in the objective or in the constraint, allows us to use the duality theorem.

In its most general form, the theorem isn’t the most useful thing going, but you will find that it provides a unifying principle that joins together most of the things we will learn this quarter (and many things you will learn beyond this quarter). The theorem also provides compelling reasons for economists to be so fond of convexity.

Consider a closed, convex set \( K \), and some point \( x \) that is not in \( K \). By virtue of the separating hyperplane theorem, we assert that there exists a hyperplane that separates \( K \) from our wayward point, \( x \). 
The separating hyperplane creates a half-space that contains $K$. If we were to take the intersection of all the half-spaces that contain $K$, then we would be left with $K$ (if $K$ were non-convex, we would only get its convex hull). So, now we have a way to relate a non-linearly parameterized convex set, such as $\{x|u(x) \geq u\}$ to linearly parameterized half-spaces. By introducing the idea of the support function of a set, we can make this relation more concrete.

Given a price vector $p$, we can take the planes orthogonal to $p$ and separate out our bundle space into regions of increasing value. So, let’s find the point in the set that gives us the lowest $p \cdot x$ value for a given $p$. The (limiting) value of this point is given by

$$\mu_K(p) = \inf\{p \cdot x| x \in K\} \quad (A.3)$$

We call this $\mu_K(p)$ the support function of the set $K$. Now we can reconstruct $K$ from its support function as described above - by intersecting the relevant half-spaces:

$$K = \bigcap_p \{x|p \cdot x \geq \mu_K(p)\} \quad (A.4)$$
So, $\mu_K(p)$ is our link from linear half-spaces to the non-linearly parameterized set $K$. Now for the duality theorem:

**Duality Theorem.** Let $K$ be a nonempty closed set and let $\mu_K(p)$ be its support function. Then, there is a unique $\bar{x} \in K$ s.t. $\bar{p} \cdot \bar{x} = \mu_K(\bar{p})$ if and only if $\mu_K(p)$ is differentiable at $\bar{p}$. What’s more, $\nabla \mu_K(\bar{p}) = \bar{x}$.

i.e. so far as derivatives of the maximizer are concerned, $\mu_K$ is linear. We also have a condition for the uniqueness of the maximizer. A little elbow grease will show that the standard conditions for uniqueness of solution in constrained optimization fit this theorem like a glove.

To illustrate, consider the EMP. The upper contour set of the utility is $\{x \in \mathbb{R}_+^L | u(x) \geq u\}$. What is its support function? Clearly, it is the expenditure, $e(p, u)$. Hence, from our theorem, we attain the envelope identity $\nabla_p e(p, u) = h(p, u)$.

Also, consider Figure A.4. Consider prices of form $(p_x, p_y) = (\frac{1}{2}, p)$. If $p < \frac{1}{2}$, then the inf of $p \cdot x$ in $K$ will occur at the point $(\frac{7}{4}, \frac{1}{4})$. If $p > \frac{1}{2}$, it will occur at $(\frac{1}{4}, \frac{7}{4})$. If $p = \frac{1}{2}$, the the inf occurs at all points on the diagonal part of the boundary.
of $K$, i.e. the arg inf becomes non-unique. So, we can define the support function by

$$
\mu_K\left(\frac{1}{2}, p\right) = p \cdot \arg \inf_{x \in K} p \cdot x = \left\{ \begin{array}{ll}
p \cdot \left(\frac{1}{4}, \frac{7}{4}\right) = \frac{1}{8} + \frac{7}{4}p & \text{for } p < \frac{1}{2} \\
p \cdot \left(\frac{7}{4}, \frac{1}{4}\right) = \frac{7}{8} + \frac{1}{4}p & \text{for } p > \frac{1}{2} \end{array} \right. 
$$

(A.5)

So, $\mu_K$ has a kink (i.e. is not differentiable) at the price vector where the arg inf is non-unique, just as predicted by the duality theorem. Additionally, note that any line with a slope from $\frac{1}{4}$ to $\frac{7}{4}$ can act as a tangent to $\mu_K$ at the kink. The duality theorem tells us that $\nabla \mu_K(p) = \arg \inf_x p \cdot x$. Since, our graph holds $p_x$ constant, its $p$ derivative tells us the $y$-coordinate of the arg inf. So, for $p < \frac{1}{2}$, the arg inf occurs at $y = \frac{7}{4}$ and for $p > \frac{1}{2}$, the arg inf occurs at $y = \frac{1}{4}$. At $p = \frac{1}{2}$, there is an arg inf for all $y \in [\frac{1}{4}, \frac{7}{4}]$. So, if $\mu_K$ is differentiable, then the slope of the tangent is unique and thus the arg inf is unique. If $\mu_K$ is not differentiable, then the slope of the tangent either doesn’t exist or is non-unique. This too tells us about the arg inf. So the duality theorem has set up a very interesting connection between the constraint space of the optimization and the support function of the optimization. You will find that duality is all around, especially when we start playing with how the EMP relates to the UMP (or how the profit maximization problem relates to the cost minimization problem).
Appendix B

Euler’s theorem and its converse

B.1 Euler’s theorem

Euler’s theorem is often taught, but rarely its converse. Both are true. This we will show.

B.1.1 $f(p)$ is homogeneous of degree 1 $\Rightarrow f(p) = p \cdot \nabla f(p)$

Define $g(\lambda) \equiv f(\lambda p)$. Then, homogeneity of degree 1 tells us

$$g(\lambda) = \lambda g(1) \quad (B.1)$$

Differentiating both sides yields

$$p \cdot \nabla f(\lambda p) = g(1) \quad (B.2)$$

Setting $\lambda = 1$ completes the proof

$$p \cdot \nabla f(p) = g(1) = f(p) \quad (B.3)$$

B.1.2 $f(p)$ is homogeneous of degree 1 $\Leftarrow f(p) = p \cdot \nabla f(p)$

Again define $g(\lambda) \equiv f(\lambda p)$. Now, if $g(1) = f(p) = p \cdot \nabla f(p)$, then

$$g(\lambda) = f(\lambda p) = \lambda p \cdot \nabla f(\lambda p) \quad (B.4)$$

From the definition of $g(\lambda)$, we see that

$$g'(\lambda) = p \cdot \nabla f(\lambda p) \quad (B.5)$$

Combining the two, we find

$$\lambda g'(\lambda) = g(\lambda) \quad (B.6)$$

This is a separable differential equation. Separating

$$\frac{dg(\lambda)}{g(\lambda)} = \frac{1}{\lambda} \quad (B.7)$$
and integrating yields
\[ \log g(\lambda) = \log \lambda + C \]  
(B.8)
or, exponentiating
\[ g(\lambda) = \lambda c \]  
(B.9)
Setting \( \lambda = 1 \), we get the initial condition
\[ g(1) = c \]  
(B.10)
So, our final solution is
\[ g(\lambda) = \lambda g(1) \]  
(B.11)
Inserting the definition of \( g(\lambda) \) completes the proof
\[ f(\lambda p) = \lambda f(p) \]  
(B.12)

\section*{B.2 Corollary to Euler’s theorem}

\subsection*{B.2.1 \( f(p) \) is homogeneous of degree 1 \( \Rightarrow \nabla f(p) \) is homogeneous of degree 0}

Define \( h(p) \equiv f(\lambda p) \). Then, homogeneity of degree one states
\[ \lambda f(p) = h(p) \]  
(B.13)
Differentiate both sides of the equation with respect to \( p \) to yield
\[ \lambda \nabla f(p) = \nabla h(p) = \lambda \nabla f(\lambda p) \]  
(B.14)
Cancelling \( \lambda \)'s completes the proof
\[ \nabla f(p) = \nabla f(\lambda p) \]  
(B.15)
Appendix C

Expansion in a basis of mins

Consider the function $g(x) = \min(x, \theta)$. It looks like

![Graph showing the function $g(x) = \min(x, \theta)$]

Now what does $g'(x)$ look like? Below $\theta$, the derivative is equal to one, and above it is equal to zero. So, loosely, $g'(x) = H(\theta - x)$, where $H(x)$ is the unit step function (0 below $x = 0$, 1 above $x = 0$).

![Graph showing the derivative $g'(x)$]

How can we show this more formally? For a given function $g(x)$, consider its action through the following functional

$$F[g(x), f(x)] = \int g(x) f(x) dx$$

(C.1)

Let $g_1(x)$ and $g_2(x)$ be considered the same function if $F[g_1(x), f(x)] = F[g_2(x), f(x)]$ for all $f(x)$. So, with $g(x) = \min'(x, \theta)$, we can expand the functional with integra-
tion by parts

\[ F[\min'(x, \theta), f(x)] = \int_0^\infty \min'(x, \theta) f(x) \, dx \]
\[ = \left. \min(x, \theta) f(x) \right|_0^\infty - \int_0^\infty \min(x, \theta) f'(x) \, dx \]
\[ = \lim_{x \to \infty} \theta f(x) - \int_0^\theta x f'(x) \, dx - \theta \int_0^\infty f'(x) \, dx \]
\[ = \lim_{x \to \infty} \theta f(x) - \int_0^\theta f(x) \, dx + \int_0^\theta f(x) \, dx - \lim_{x \to \infty} \theta f(x) + \theta f(0) \]
\[ = \int_0^\theta f(x) \, dx \]
\[ = \int_0^\infty H(\theta - x) f(x) \, dx \]

\[ F[\min'(x, \theta), f(x)] = F[H(\theta - x), f(x)] \quad (C.2) \]

So, indeed, if we use this functional notion of equivalence, then \( \min'(x, \theta) = H(\theta - x) \).

Going even further, what is the derivative of \( H(\theta - x) \), i.e. what is \( \min''(x, \theta) \)?

\[ F[\min''(x, \theta), f(x)] = F[H'(\theta - x), f(x)] \]
\[ = \int_0^\infty H'(\theta - x) f(x) \, dx \]
\[ = H(\theta - x) f(x) \left|_0^\infty \right. - \int_0^\infty H(\theta - x) f'(x) \, dx \]
\[ = -f(0) - \int_0^\theta f'(x) \, dx \]
\[ = -f(0) - f(x) \big|_0^\theta \]
\[ = -f(0) - f(\theta) + f(0) \]
\[ = -f(\theta) \]
\[ F[\min''(x, \theta), f(x)] = F[-\delta(x - \theta), f(x)] \quad (C.3) \]

where the Dirac \( \delta \) in the last line is defined by the functional relation \( f(\theta) = \int_0^\infty \delta(x - \theta) f(x) \, dx \).

Now, if \( \min''(x, \theta) \) is the negative Dirac \( \delta \), as defined above, then for any function, \( f(x) \), we can write it equivalently as \( -\int_0^\infty f(\theta) \min''(x, \theta) d\theta \). What if we then use integration by parts to pull the derivatives from the min to the \( f \)? Then, we find

\[ -\int_0^\infty f(\theta) \min''(x, \theta) d\theta = -f(\theta) \min'(x, \theta) \big|_0^\infty + \int_0^\infty f'(\theta) \min'(x, \theta) d\theta \]
\[ = -f(\theta) H(\theta - x) \big|_0^\infty + \int_0^\infty f'(\theta) \min(x, \theta) \big|_0^\infty - \int_0^\infty f''(\theta) \min(x, \theta) d\theta \]
\[ = f(0) + x \lim_{\theta \to \infty} f'(\theta) - \int_0^\infty f''(\theta) \min(x, \theta) d\theta \quad (C.4) \]

But the left hand-side of this expansion is equal to \( f(x) \). So, this is just an expansion of \( f(x) \) on the basis \( \{1, x, \min(x, \theta)\} \).

\[ f(x) = f(0) + x \lim_{\theta \to \infty} f'(\theta) - \int_0^\infty f''(\theta) \min(x, \theta) d\theta \quad (C.5) \]

For this expansion to converge, we just need for the \( \lim_{\theta \to \infty} f'(\theta) \) to not explode. A necessary condition for this is that \( f \) is everywhere increasing and concave. Hence,
any utility function with the standard assumptions can be expressed as a min series. In this sense, functions of the form $\min(x, \theta)$ are the canonical concave functions in the same way that sines and cosines are the canonical periodic functions. Thus, any time you are called to construct an example with a concave function, know that a min is a good place to start.
Appendix D

Assortment in continuous general equilibrium

Say we have some pairing scheme \( g : [0, A] \mapsto [A, 1] \) such that every worker \( a \in [0, A] \) is matched with manager \( g(a) \in [A, 1] \). What’s more, let \( g(a) \) be one-to-one and onto, and let \( \mu \) measure sets relative to the PDF \( f(a) \).

Moreover, let \( \mu([0, A]) = \mu([A, 1]) = \frac{1}{2} \) so that the support of \( f(a) \) is \([0, 1]\) and there are an equal number of managers and workers. If we know that \( g \) is decreasing (i.e. that workers and managers are substitutes), then what must \( g \) be?

We derive this assuming that the mapping is continuous, so that a small chunk of workers will match to a small chunk of managers. Using differentials, we then find that if \( a_M = g(a_W) \), then the chunk of workers \([a_W, a_W + da_W]\) maps to the chunk of managers \([a_M + da_M, a_M]\) where

\[
da_M = g'(a_W)da_W \tag{D.1}
\]

Note that \( da_M \) is of the opposite sign than \( da_W \) since \( g(a) \) is a decreasing function (i.e. we are negatively assortative). This simply maps regions of talent to each other. Then, we will need to add the further constraint that the measure of workers in \( da_M \) is equal to the measure of workers in \( da_W \). Hence

\[
\begin{align*}
\mu([a_M + da_M, a_M]) &= \mu([a_W, a_W + da_W]) \\
|da_M|f(a_M) &= |da_W|f(a_W) \\
-f(a_M)g'(a_W)da_W &= f(a_W)da_W \\
-f(a_M)g'(a_W) &= f(a_W)
\end{align*} \tag{D.2}
\]

And, since \( a_M = g(a_W) \), we come to

\[
-f(g(a_W))g'(a_W) = f(a_W) \tag{D.3}
\]

This is a separable differential equation in \( g(a) \) and \( a \). Solving

\[
\begin{align*}
-f(g(a))dg(a) &= f(a)da \\
\int g'(a)f(\gamma)d\gamma &= \int_0^a f(\alpha)d\alpha \\
1 - F(g(a)) &= F(a) + C \\
g(a) &= F^{-1}[1 - F(a) - C]
\end{align*} \tag{D.4}
\]
Now, we just have to use initial conditions to get $C$. In a negatively assortative scheme, we must have the median of the distribution be a fixed point (i.e. the chunk just below the median must match with the chunk just above the median). So, let the median $a_{med}$ be implicitly determined by

$$F(a_{med}) = \int_0^{a_{med}} f(\alpha)d\alpha = \frac{1}{2} \tag{D.5}$$

Then, the constant $C$ is determined by solving

$$g(a_{med}) = F^{-1}[1 - F(a_{med}) - C] = a_{med}$$

$$1 - F(a_{med}) - C = F(a_{med})$$

$$1 - \frac{1}{2} - C = \frac{1}{2}$$

$$C = 0 \tag{D.6}$$

So we have derived a formula for the only continuous assortment scheme that is negatively assortative.

$$g(a) = F^{-1}[1 - F(a)] \tag{D.7}$$