Natural Resources as Capital: Theory and Policy[1]

Larry Karp

[1] This manuscript benefitted from comments by Sangeeta Bansal, Edward Barbier, Stephen Holland,Tobias Larsen, Karolina Ryszka., Jesus Marin Solano, Zhen Sun, Yang Xie, and Huayong Zhi, and research assistance of Wenfeng Qiu. It also benefitted from the test-runs provided by the 2013 –2015 classes in EEP/Econ 102.

Contents

Preface

Please send suggestions and corrections to: karp@berkeley.edu. For the online version of this book, click here.

The book is designed for upper division undergraduates. Together with the appendices, it is also suitable for a masters level course. Prerequisites include an intermediate micro-economics course and a grounding in calculus. The presentation uses derivatives, and in a few cases partial and total derivatives. Appendix B reviews the required mathematical tools. Asterisks identify sections with more advanced material.

The text covers standard resource economics topics, including the Hotelling model for nonrenewable resources, and renewable resource models such as fisheries. The distinction between natural resource and environmental economics has blurred and become less useful over the decades. This book reflects that evolution by including some topics that also fit in an environmental economics text, while still emphasizing natural resource topics. For example, the problem of climate change involves resource stocks, and therefore falls under the rubric of natural resources. Environmental externalities drive the problem, so the topic also fits in an environmental economics text.

Two themes run through this book. First, resources are a type of natural capital; their management is an investment problem, requiring forward-looking behavior, and thus requiring dynamics. Second, our interest in natural resources stems largely from the prevalence of market failures, notably incomplete or nonexistent property rights. "Policy failures" complicate matters; in many circumstances, policy is inadequate to address market failures, or exacerbates those failures. The book emphasizes skills and intuition needed to think sensibly about dynamic models, and about regulation in the presence of both market and policy failures. The opportunity cost of this focus is the omission of a detailed discussion of several important resources (e.g. forestry). This pedagogic decision reflects the view that upper division (and masters-level) students are better served by acquiring a good understanding of concepts and tools that will help them to think critically about a broad range of resource issues. Students at this level are already adept at acquiring information about whatever subject interests them. It is more difficult to master (or even identify) the concepts and tools that promote critical analysis of resource issues.

Standard topics The eleven chapters, (3 – 5, 7, and 12 – 18 cover the nuts and bolts of resource economics, and could stand alone as a mini-course. Chapters 3 – 5 and 7 cover nonrenewable resources. Chapters 3 and 4 study the two-period model, first in the simplest setting and then including stock-dependent costs. Chapter 4 explains the role of resource scarcity and stock-dependent extraction costs in determining resource rent. This chapter shows how to obtain the optimality condition using the "perturbation method"; Chapter 5 adapts that method to the $T-$period setting to obtain the Euler equation, known in this context as the Hotelling condition. The perturbation method (the discrete time calculus of variations) enables students to perform constrained optimization almost without being aware of it: it is simpler and more intuitive than the method of Lagrange. We use the two-period optimality condition to write the T-period optimality condition merely by replacing time subscripts. Chapter 5 also discusses the idea of the shadow value of a resource stock, and illustrates the transversality condition by means of an example. Chapter 7 presents the backstop model. Backstops are economically important, and this material gives students practice working with the Hotelling model while preparing for a subsequent policy-focused chapter.

Chapters 12 – 18 study renewable resources. We emphasize fisheries because these provide a concrete setting, and they illustrate most of the issues found in other renewable resources. Chapter 12 defines and provides historical perspective on different types of property rights, and then discusses the Coase Theorem. The rest of the chapter uses real world examples and a one-period analytic model to describe the difficulties and the unintended consequences arising from fishery regulation. It discusses attempts to establish property rights in fisheries, as an alternative to regulation.

Chapter 13 introduces the concepts needed to study renewable resources, including the growth function, steady states, (local) stability, and maximum sustainable yield. This chapter explains the relation between and the relative advantages of discrete time versus continuous time models, and describes how the rest of the text uses these two approaches. Chapter 14 discusses the open access fishery, showing how the long run effect of policy depends on the initial stock size. Chapter 15 introduces the sole owner fishery, deriving the Euler equation and discussing policy under multiple market failures. We compare steady states under open access and for the sole owner. We explain the effect of harvest cost and the discount rate on the sole owner's steady state(s). Chapter 16 shows how to analyze the sole owner fishery outside the steady state. There we begin with a problem that can be studied using only the Euler equation and careful reasoning; we then move to a more complicated example requiring phase portrait analysis. Chapter 17 discusses water economics, showing how the tools developed for the fishery setting can be adapted to other resources. Chapter 18 explains concepts of weak and strong sustainability and then discusses the Hartwick rule. Concern over sustainability has led to the development of modifications and alternatives to gross national product (GNP) as measures of welfare.

Less standard topics Chapter 2 reviews topics in micro-economics needed to study resource economics. These topics include the concept of arbitrage, the use of elasticities, the relation between competitive and monopoly equilibria, and the use of discounting. Chapter 6 discusses empirical tests of the Hotelling model.

Chapters 8 – 11 introduce policy problems. Chapter 8 uses the Hotelling model to examine the "Green Paradox", an important topic in climate policy. In addition to its intrinsic interest, this material gives students practice in using the Hotelling model, and more generally illustrates the use of models to study policy questions. The material promotes critical thinking by discussing limitations of the green paradox model.

Chapter 9 provides the foundation for policy analysis when market failures are important. It explains and illustrates the Theory of the Second Best and the Principle of Targeting, and discusses the importance of political lobbying and the distinction between policy complements and substitutes. In order to present this material simply, examples in the chapter use static environmental problems, instead of stock-dependent natural resource problems. Chapter 12 uses the concepts developed here, again in a one-period setting. Chapter 14 then develops these concepts in a dynamic setting.

Taxes and other market-based instruments are becoming increasingly important regulatory tools. Chapter 10 introduces the principles of taxation in a static framework. We discuss tax equivalence, tax incidence, and deadweight loss. These basic ideas provide a conceptual framework for estimating the fraction of permits in a cap and trade scheme that would need to be grandfathered in order to compensate firms for the cost of regulation. Chapter 10 is essential for understanding taxes in the dynamic natural resource setting, the topic of Chapter 11. That chapter provides an overview of actual taxation (and subsidy) or fossil fuels, and then explains how to synthesize the Hotelling model with the information on taxes studied in the static setting. A numerical example illustrates this synthesis.

Chapter 19 studies the role of discounting, emphasizing its role in recommendations for optimal climate policy. The chapter explains the difference between utility and consumption discounting. It discusses the Ramsey formula for the social discount rate in the deterministic setting, and then introduces uncertainty. We emphasize the importance, to the social discount rate, of projections of future economic growth. A concluding section introduces hyperbolic discounting and explains its relevance to climate policy.

Chapter 1

Resource economics in the Anthropocene

Natural resources are under threat of misuse and depletion, but human ingenuity makes it possible to devise rules and create institutions to protect them. Policies that harness the power of markets are more likely to be successful. Resource economics offers a framework for analyzing resource use, providing tools that can contribute to improved stewardship.

Natural resources are a type of capital: natural, as opposed to man-made capital. Resource use potentially alters the stock of this capital, and is a type of investment decision. Change is thus a key feature of natural resources, requiring a dynamic (i.e., multiperiod) perspective. Natural resources, like other types of capital, provide services that affect human well-being. In some cases, as with burning oil or eating fish, we consume those services by consuming a part of the resource. In other cases, we consume natural resource services indirectly. Wetlands provide filtration services, reducing the cost of clean water. Transforming wetlands into farms or cities changes the flow of these services. Other resource stocks, such as bees and bats, critical to agricultural production, provide indirect but essential services. Diminishing those stocks, by reducing habitat or otherwise changing the ecosystem, alters future pollination services.

These examples are anthropocentric, attributing value to natural resources only because they provide services to humans. Species or wilderness areas may have intrinsic value apart from any effect, however indirect, they have on current or future human welfare. Regardless of whether one begins with a purely anthropocentric view or a more spiritual/philosophical perspective, resource value depends on the remaining stocks: oil, fish, wetlands, pollinators, wilderness etc. These stocks affect current and future flows of services. Spiritual or philosophical arguments can be effective in galvanizing public action, but the anthropocentric view, putting humans at the center of the narrative, can lead to more effective remedies. The protection of forests or fisheries is more likely achieved if the people near these resources have a stake in their protection. Ecotourism in nature preserves can give local residents an incentive to respect the preserve; using the dung of elephants and rhinos to create paper products for sale abroad gives people an incentive to protect the animals when they stray from the preserves. Elevating philosophical abstractions above concrete human needs risks promoting inefficient policy.

In common usage, "capital" refers to man-made productive inputs, such as machinery, or the monetary value of those inputs. A broader definition treats capital as anything that yields a flow of services. Education augments our stock of human capital, making us more productive or otherwise enhancing our lives. Natural resources fall under this broader definition of capital. A firm's decision about purchasing additional machinery, or an individual's decision about acquiring more education, are investment decisions, and thus "forward looking"; they depend on beliefs about their future consequences. The decisions are "dynamic" rather than "static" because their temporal aspect is central to the decision-making process.

Natural capital, like machinery, obeys laws of physics: trees age, machines rust. The fact that the two types of capital inhabit the same physical universe connects resource economics to the broader field of economics. However, natural and man-made capital differ in the severity of the market failures that affect them. An "externality", a type of market failure, arises when a person does not take into account all of the consequences of their action. Unregulated pollution or excessive use of a resource stock are leading examples of externalities. Externalities and other market failures, often associated with weak or nonexistent property rights, are *central* to the study of resource economics.

A natural resource without property rights cannot be bought and sold, and therefore does not have a market price reflecting its value. Informed policy decisions require a comparison of costs and benefits of different alternatives, e.g. protecting the natural resource or allowing development. The lack of a price for the natural resources makes it difficult to value its services, greatly complicating the policy problem. (Box 1.1).

Changes in resource stocks occur either intentionally, via the exercise of property rights, or accidently, arising from externalities. A mine owner understands that extracting a unit of ore today means that this material cannot be sold in the future. Where property rights to resources do not exist, individuals often make decisions without regard to their effect on resource stocks. Farmers choose levels of pesticide and fertilizer in order to increase their profits. Some of these inputs enter waterways, where they damage the publicly owned ecosystem, as has occurred in the Everglades and in the Gulf of Mexico. Individual farms have negligible effect on the aggregate outcomes, and individual farmers have no (selfish) interest in those outcomes, so it is rational for them to ignore these consequences. In these cases, society chooses how to use publicly owned resources. Often the choice, made by default, involves little resource protection.

Box 1.1 Valuing natural capital. (a) The cost of protecting watershed-based filtration systems for New York City's water supplies was estimated in 1996 at \$1–1.5 billion; the cost of building and operating a filtration system was estimated at \$6–8 billion. New York City protected the watershed. (b) The value of irrigation in a region of Nigeria was 4–17% of the losses to downstream floodplains arising from the diminished water flows caused by this irrigation. (c) Shrimp farming in Thailand causes the destruction of mangrove swamps, which provided nurseries for other fisheries and storm protection. The value of shrimp farming was about 10% of the lost value of ecosystem services. The estimates for (b) and (c) were made after the damage had occurred (the irrigation was put in place, and the shrimp farms developed). It may be costly or politically infeasible to undo these actions, e.g. to reduce irrigation in order to increase water flows, or to restore the mangroves.

Rapid changes in resource stocks, and the expected change of future stocks, make resource economics an especially important field of study. The 2005 Millennium Ecosystem Assessment reports that the recent speed and extent of human-induced change in ecosystems is greater than the world has previously seen. Some changes, such as those associated with the expansion of agricultural production, contribute to current human well-being. However, many of the changes degrade resources, eroding natural capital and threatening future well-being. The last half-century saw a fifth of the world's coral reefs lost, and another fifth degraded; a third of mangrove forests have been lost; water use from rivers and lakes has doubled; nitrogen flows entering ecosystems have doubled, and phosphorus flows have tripled. Human actions have increased the rate of species extinction as much as 1000-fold, relative to rates found in the geological record. Estimates of mammal, bird and amphibian species threatened with extinction range from $10-30\%$. Tropical forests and many fisheries are in decline.

The Millennium Assessment evaluates 24 different types of ecosystem services and concludes that 60% of these are degraded or threatened. The loss in natural capital may lead to abrupt changes, e.g. flips in water quality (eutrophication) and the rapid emergence of new diseases. These damages tend to disproportionately harm the poor and most vulnerable. Barring major policy changes or technological developments, ecosystems will likely face increasing pressure. Standard measures of wealth ignore these changes in natural capital. Attempts to account for natural capital, show that almost half of countries in a World Bank study are depleting their wealth, living off natural capital.

Climate change may pose the single greatest danger to future well-being. Climate change is likely to exacerbate the types of problems already seen, such as loss of species, the spread of diseases, and increased water shortages. It may also lead to new problems, including rising sea levels, increased frequency of severe weather events, and decreased agricultural productivity. The costs of these changes will depend on uncertain relations between stocks of greenhouse gasses and changes in temperature, ocean acidity, and the sea-level, and on the uncertain relation between these variables and economic and ecological consequences (e.g. decreased agricultural productivity and species loss). The future stocks of greenhouse gasses depend on future emissions, which depend on uncertain changes in policy and technology.

In recognition of man's ability to fundamentally alter the earth's ecosystem, many scientists refer to the current geological period as the Epoch of the Anthropocene ("New Human Epoch"). Proposals for this Epoch's starting date range from the early industrial age to the middle of the 20th century.

The view that current resource use will create large costs to future generations leads to resource pessimism. Thomas Malthus, an early resource pessimist, claimed that if unchecked by war, disease, or starvation, human population tends to rise faster than food production. He concluded that population eventually outstrips food supplies, until starvation, war, or disease brings them back into balance. This description was quite accurate for most of human history, but events since he wrote in 1800 have contradicted his predictions. Many countries have seen a demographic shift associated with higher income, leading to stabilization or decreases, not increases, in population. In poor societies, children are a form of investment for old age. In rich societies, children do not provide the primary support for their aged parents, and the cost of raising children is high. These factors encourage smaller family sizes with rising income. Technological innovations have increased agricultural productivity and reduced the cost of transporting and storing food. Population and food security have both increased. Most recent famines were caused not by the absolute lack of food, but by its unequal distribution.

In the 19th century the British government was concerned that high consumption of coal would lead to future scarcity. William Jevons, a prominent economist at the time, advised the government not to use policies that would lead to coal conservation, on the ground that the market would resolve any future problem: if the price of coal did rise, businesses would reduce their demand, and innovators would develop substitutes for coal. In 1931 Harold Hotelling produced one of the cornerstones of the field of resource economics, responding to the pessimists of his time. He wrote

> Contemplation of the world's disappearing supplies of ... exhaustible assets has led to demands for regulation of their exploitation. The feeling that these products are now too cheap for the good of future generations, that they are being selfishly exploited at too rapid a rate, and that in consequence of their excessive cheapness they are being produced and consumed wastefully has given rise to the conservation movement.

Hotelling studied the use of natural resources in an idealized market with perfect property rights, where a rational owner takes the finite resource supply into account. In this setting, prices signal scarcity, influencing decisions about extraction, exploration, and the development of alternate energy sources and new technologies. Prices signals can also lead to fundamental changes in human behavior, such as family size.

Barnett and Morse (1963) examined trends in resource prices, finding no evidence of increased scarcity. However, increased resource use led to a resurgence in resource pessimism, exemplified by Paul Ehrlich's *The Population Bomb.* Ehrlich, a biologist, observing rapid increases in the use of natural resources in the 1960s, and accustomed to working with mechanistic models of insect populations, predicted imminent and catastrophic resource scarcity. Julian Simon, like Jevons almost a century earlier, thought that the market would take care of scarcity, as higher prices encouraged exploration, discovery, production and conservation. Simon proposed, and Ehrlich accepted, a bet that the inflation-adjusted price of a basket of five minerals would fall over a decade. This period of time seemed long enough to test Ehrlich's forecast of imminent scarcity. Simon won the bet; Ehrlich claimed that he had underestimated the rapacity of man's resource extraction, and that his prediction of scarcity was wrong only in the timing. (Appendix A)

Box 1.2 The (im)possibility of extinction. In the mid 1800s, driftnet herring fishermen asked for regulation to restrict the use of "longlines", which they claimed damaged fish stocks and reduced catches. Many scientists, believing that the self-correcting power of nature would take care of any temporary problems, resisted those requests. The influential scientific philosopher Thomas Henry Huxley, a member of British fishing commissions charged with investigating the complaints, explained in 1883 why the requests were unscientific, and merely designed to impede technological progress: "Any tendency to over-fishing will meet with its natural check in the diminution of the supply,... this check will always come into operation long before anything like permanent exhaustion has occurred."
Others disagreed. Maine's fishery commissioner Edwin Gould stated in 1892 "It's the same old story. The buffalo is gone; the whale is disappearing; the seal fishery is threatened with destruction. Fish need protection."

The resource pessimism of the 1960's led to renewed interest in resource economics during the 1970s and 1980s. The dominant strand of this literature extends Hotelling's earlier work, using the paradigm of rational agents operating with secure property rights. However, there has also been increased recognition of market failures, especially externalities associated with missing markets and weak property rights. Modern resource economics provides a powerful lens through which to study natural resources precisely because it takes market failures seriously. The discipline provides a counterweight to pessimists' tendency to understate society's ability to respond to market signals, while also providing a remedy to the excessively optimistic belief that markets, by themselves, will solve resource problems.

Agriculture and fisheries illustrate the power of markets and the problems arising from market imperfections. In both cases, markets have unleashed productivity gains leading to abundance. But these gains occur in the presence of market failures, threatening (in the case of agriculture) or reversing (in the case of fisheries) the initial gains. Neither markets nor natural forces will automatically solve these problems without policy intervention. Increases in agricultural productivity since the 1960s made it possible to feed twice the population with slightly more than a 10% increase in farmed land, reducing or eliminating the threat of starvation for hundreds of millions of people. Those changes were associated with increased pesticide and fertilizer use that threaten waterways, increased and likely unsustainable use of water, and increased loss of habitat. Rising fish harvests reduced fish stocks, resulting in relative scarcity and higher prices. Responding to market signals, fishers adopted new technologies, increasing their ability to catch fish. These gains have often been short-lived, as the increased harvest degrades fish stocks, ultimately lowering harvest.

Markets have been essential in "disproving" the resource-pessimists thus far. Markets are powerful in part because they are self-organizing. They require a legal and institutional framework that respects private property and contracts; they often require regulation, but not detailed governmental management. However, the beneficial changes assisted by markets, occurring in the context of market imperfections, may in the longer run validate the resource-pessimists. Where market imperfections are severe, markets are unlikely to solve, and may exacerbate resource and environmental problems.

There are objective technological and demographic obstacles to solving resource problems, but politics also create obstacles. Proposed remedies usually create winners and losers, with the losers often in a better position to defend their interests. For example, effective climate policy will reduce fossil fuel owners' wealth; this group is politically powerful.

The resource policies actually in use emerge in the political marketplace, both in democracies and under other forms of governance. Some policies are driven by self-interest and not explicitly linked to resource issues, while still having direct and harmful effects on resources. In a few cases, there is a near-consensus (at least amongst economists) that the policies harm natural capital and more broadly are socially irrational. Prominent examples include: agricultural policies that promote environmentally damaging production along with commodity gluts; water policy that promotes excessive use and inefficient allocation across users; fossil fuel subsidies that exacerbate the problem of excessive greenhouse gas emissions; fishing subsidies that worsen the problem of overharvest. Other policies, such as the promotion of corn-based ethanol by the US Renewable Fuel Standard, are at best questionable. Still other policies, such as US fishing regulation during the last two decades, improve on previous policies, but still fall short of achieving their objectives.

These are all examples of policy failure, some naked and some nuanced, some extreme and others mild. The worst policies can be explained by political power in the service of self-interest. The inherent difficulty of managing complex problems even where there is good will also explains policy limitations. Market failures require a policy response, but experience shows that policy intervention sometimes is part of the problem, not part of the solution.

Clearer thinking will not dispel the technical, demographic and political obstacles to socially rational resource use. However, clearer thinking and more precise language can help overturn prejudice and identify effective policy, and can provide a basis for negotiations. People might disagree on a conservation measure, but it is counter-productive to base the disagreement on identification with a political party or a disciplinary speciality (economics versus ecology). Resource economics can provide a common language and analytic framework, creating the possibility of moving beyond ideology.

Resource economics also helps in understanding that institutional reform, such as the creation of property rights rather than the introduction of a new tax, is often an effective remedy to problems. Some people distrust property rights because they (correctly) see these as the basis for markets, and they (probably incorrectly) think that markets are responsible for the resource problem. Resource economics teaches that many problems are due not to markets, but to market *failures*.

Disagreements about resource-based problems tend to be easier to resolve where the problems are local or national rather than global, and where changes occur quickly (but not irreversibly), rather than unfolding slowly. The local or national context makes the horse-trading needed to compensate losers easier. The rapid speed of change makes the problem more obvious, and makes the potential benefits of remedying the problem, and the costs of failing to do so, more pressing for the people who need to engage in this horse-trading. The most serious contemporary resource-based problems are global and unfold over long periods of time, relative to the political cycle.

A prominent international treaty, the Montreal Protocol, helped to reverse the global problem of ozone depletion. The rapid increase in the ozone hole over the southern hemisphere made the problem hard to ignore, and the availability of low-cost alternatives to ozone-depleting substances made it fairly cheap to fix. International negotiations on other global problems, prominently climate change, have been notably unsuccessful. It is difficult to summon the political will to make the international transfers needed to compensate nations that would be, or think they would be, better off without an agreement. The most serious effects of climate change will impact future generations, who have no direct representation in current negotiations.

An overview of the book

Two points made above set the stage for the rest of this book. (i) Markets have the potential to ease environmental and resource constraints, contributing hugely to the increase in human welfare. (ii) Many problems arise from market failures, such as externalities associated with pollution; those market failures may diminish or even reverse the beneficial effects arising from markets that function well. A corollary to these claims is that regulation that harnesses the power of market forces, or the establishment of property rights, may make it easier to solve resource problems. Those regulations and institutional changes require political intervention; they do not arise spontaneously from market forces. This book develops and uses a theoretical apparatus that can contribute to coherent analysis of these issues. Theory makes it possible to intelligently evaluate the facts of specific cases, in pursuit of better policy prescriptions.

The pedagogic challenge arises because resources are a type of capital, requiring a dynamic setting in which agents are *forward looking.* In a static setting, firms' and regulators' decisions depend on current prices and (for example) pollution. In the dynamic resource setting, a firm's decision on how much of the resource to extract and sell in a period depend on the price in that period, and the firm's beliefs about future prices. A regulator's (optimal) policy depends on beliefs about future actions, and these depend on future prices. This difference between the static and dynamic setting is central in our presentation of resource economics.

The first half of the book provides the foundation for studying nonrenewable resources, such as coal or oil. This foundation requires a review of some aspects of microeconomic theory, and the development of methods needed to study dynamic markets. We apply these methods to the resource problem, emphasizing perfectly competitive markets with no externalities or other distortions. This material identifies incentives that are important in determining resource use, and helps the reader understand the potential for markets, when they work well. The second half of the book applies these tools, emphasizing situations where market failures create a rationale for policy intervention. We consider both the possibility that policies ameliorate the market failure, and the possibility that policy is harmful due to unintended consequences. We also move from nonrenewable to *renewable* resources, making it possible to show how different systems of property rights and policies alter resource levels at different time scales.

Terms and concepts

Epoch of the Anthropocene, renewable versus nonrenewable resource, eutrophication, market failure, externality, resource pessimist/optimist

Sources

Barnett and Morse (1963), observing that inflation-adjusted resource prices were not trending upward, concluded that there was no evidence of increased resource scarcity.

The United Nations' (2005) *Millennium Ecosytem Assessment* reports an international group of scholars' assessment of recent environmental changes, their consequences on human well-being, and likely scenarios for future changes.

Alix-Garcia et al (2009) discuss the payment of environmental services in agriculture, an example of a market-based remedy to externalities.

Elizabeth Kolbert *The Sixth Extinction: an Unnatural History* documents species extinction.

Duncan Foley (2006) *Adam's Fallacy* discusses Thomas Malthus and other important economists.

Scott Barrett *Environment and Statescraft* (2003) discusses the difficulty of creating effective global environmental agreements.

Gregory Clark *A Farewell to Alms* provides a long run historical perspective on Mathus' ideas.

Paul Sabin, "*The Bet*" examines the tension between resource optimists and pessimists, in particular between Ehrlich and Simon.

The World Bank's (2014) *Little Green Data Book* provides statistics on green national accounting.

Robert Solow (1974) provides the quote from Hotelling (1931).

Partha Dasgupta (2001) *Human Well-Being and the Natural Environment* develops the concept of natural resources as a type of capital.

Edward Barbier *Capitalizing on Nature: Ecosystems as Natural Assets* extends this concept to ecosystems, or ecological capital, e.g. wetlands, forests, and watersheds. The examples in Box 1 are taken from his book.

The quote from Huxley in Box 1.2 is from Kurlansky (1998), and the quote from Gould is from Bolster (2015).

Chapter 2

Preliminaries

Objectives

- Prepare students to study resource economics.

Information and skills

- Understand the meaning of arbitrage, the distinction between exogenous and endogenous variables and the use of comparative statics.
- Be able to calculate and know the definition and purpose of elasticities.
- Understand the relation between a competitive and a monopoly outcome.
- Know the definition and purpose of a discount rate and a discount factor; use them to calculate present values.
- Know the basics of welfare economics, in particular the fact that in the absence of market failures, a competitive equilibrium is efficient.

This chapter reviews and supplements the micro-economic foundation needed for natural resource economics. In the familiar static setting, a competitive equilibrium occurs when many price-taking firms choose output to maximize their profits. A competitive equilibrium in the resource setting involves a time-path of output. Resource-owning firms begin with an initial stock of the natural resource, and decide how much to supply in (typically) many periods, not just in a single period. For a non-renewable resource such as coal or oil, the cumulative extraction over the firm's planning horizon cannot exceed the firm's initial stock level. For renewable resources such as fish, natural growth can offset harvest, causing the stock to either rise or fall over time. Discounting makes profits in different periods commensurable, and is key in framing the firms' optimization problems.

"Arbitrage" takes advantage of differences, often price differences, in different markets. Moving a commodity from one place to another involves "spatial arbitrage". The resource firm engages in "intertemporal arbitrage", in deciding to sell a unit of resource in one period instead of another. Intertemporal arbitrage is the basis for understanding equilibria in resource markets, and spatial arbitrage, discussed in this chapter, provides the foundation for understanding intertemporal arbitrage. The two types of arbitrage can be studied using similar methods.

Models can help in understanding how changes in data, or an assumption, affect an "endogenous" outcome. The data/assumption is "exogenous"; it is taken as given, i.e. determined "outside " the model. For example, we might take the cost of shipping goods as exogenous, and ask how a change in this cost alters the endogenous quantity shipped, and ensuing price. Answering this type of question uses comparative statics. We sometimes use elasticities for comparative statics questions. Elasticities provide a unit free measure of the relation between two variables, such as quantity and price; "unit free" means, for example, that the relation does not depend on whether we measure prices in dollars per pound or Euros per kilo.

We emphasize competitive equilibria. However, many important resource markets, including markets for petroleum, diamonds, and aluminium (produced using bauxite, a natural resource) are not, or have not always been, competitive. We therefore supplement the study of competitive markets by considering the case of monopoly. Few, if any resource markets are literally monopolistic, but the monopoly model provides a limiting case, against which to compare perfect competition. Many resource markets lie somewhere on the continuum between these two. We use the elasticity of demand to relate the equilibrium conditions under perfect competition and monopoly.

Two welfare theorems explain the circumstances under which a profit maximizing competitive industry and a welfare maximizing social planner lead to the same outcome. These theorems also provide the basis for studying market failures.

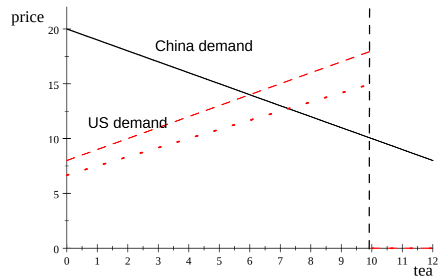

Figure 2.1: Horizontal axis shows *tea sold in China* Solid line shows demand for tea in China. Dashed line shows demand for tea in US. Dotted line shows price received by Chinese exporter with 30% transportation costs.

2.1 Arbitrage

Objectives and skills

- Understand arbitrage and graphically represent and analyze the "no-arbitrage" condition.

Much of the intuition for later results rests on the idea of "arbitrage over time". This idea is closely related to the more familiar idea of arbitrage over space. Suppose that there are ten units of tea in China, where the inverse demand is $p^{\text{China}} = 20 - q^{\text{China}}$; p^{China} and q^{China} are price and quantity consumed in China. A demand function gives quantity as a function of price, and the *inverse demand function* gives price as a function of quantity. The inverse demand for tea in the U.S. is $p^{\text{U.S.}} = 18 - q^{\text{U.S.}}$.

Figure 2.1 shows the inverse demand in China (solid line) and in the U.S. (dashed line). Moving left to right on the horizontal axis increases consumption in China, and decrease U.S. consumption, because total supply is fixed at ten. The U.S. demand function is therefore read "right to left"; the point 3 on the tea axis means that China consumes 3 units and the U.S. consumes 7 units. Except where noted otherwise, we assume that the equilibrium is "interior", meaning (here) that some tea is sold in both countries.

If transportation is free, and sales are positive in both countries, then price in a competitive equilibrium must be equal in the two countries. This "no-arbitrage" condition is necessary for profit maximization: if it did not hold, a price-taking trader could buy a unit of tea in the cheap country and sell it in the expensive country, increasing profits. The no-arbitrage (= profit maximizing) condition implies that China consumes 6 units, the U.S. consumes 4 units, and the equilibrium price, 14, is the same in both countries.

It is easy to become confused about causation in a competitive equilibrium. The trader in our tea example *does not* move tea from one location to another in order to cause the prices in the two locations to be the same. This competitive trader *takes prices as given*, and continues to move tea from one location to another until no further trade is profitable. The cross-country equality of price is a consequence, not the purpose, of trade.

Transportation costs are important in the real world. These costs can be expressed on a per unit basis (some number of dollars per unit) or on an ad valorem basis (some percent of the value). In the latter case, costs are sometimes called "iceberg costs"; it is as if a certain fraction of the value melts in moving the good from one place to another. These costs include the physical cost of transportation, and ancillary costs of setting up distribution networks, acquiring information about prices in different locations, and insurance. Denote the iceberg cost as b (for "berg"); expressed as a percent, the cost is $b \times 100\%$. For $b = 0.3$, transportation cost equals 30% of the purchase price. An exporter who buys the good for p^{China} and spends bp^{China} to transport the good, has a total unit cost of $(1+b)\, p^{\text{China}}$. In a competitive equilibrium with positive sales in both countries, $(1+b)\, p^{\text{China}} = p^{\text{U.S.}}$, or

$$p^{\text{China}} = \frac{1}{(1+b)} p^{\text{U.S.}}. \tag{2.1}$$

Equation 2.1 is the no-arbitrage condition in the presence of iceberg transportation costs, $b \geq 0$. If $b > 0$, then equation 2.1 implies that the U.S. price exceeds the China price in a competitive equilibrium.

The dotted line in Figure 2.1 shows the U.S. demand, adjusted for 30% transport costs ($b = 0.3$). A point on this dotted line gives the amount that an exporter receives, at a given level of U.S. consumption, after the exporter pays transportation costs. The equilibrium sales occurs at the intersection of the solid line and this dotted line, where the price in China equals the price, net of transportation costs, that an exporter receives for U.S. sales. In equilibrium China consumes 7.8 units, at price 12.2; the U.S. consumes the remaining 2.2 units. The U.S. transportation-inclusive price, a point on the U.S. demand function (the dashed line), is $1.3 \times 12.2 = 15.9$. Net of transportation costs, the exporter receives 12.2, a point on the dotted line.

Here, "the law of one price" holds: the price, adjusted for transport costs, is the same in both locations. This "law" describes a tendency. Arbitrage requires that people have information about prices in different locations. If this information is imperfect, then arbitrage creates a tendency for prices to move together, but not price equality. The potential to gain from price differences gives people an incentive to acquire information, but the individual selling tea in China need not know the U.S. price. If a chain of people in markets between China and the U.S. each knows only the price in neighboring markets, they know whether it is profitable to move tea east or west. Markets "aggregate" information; the middlemen move the commodity from where the price is low to where it is high, in the process revealing information about and also reducing price differences. Technological advances make the flow of information cheaper and reduce transportation costs, assisting the forces of arbitrage. Apps make it possible to instantly compare prices in different stores. Farmers in developing countries use cell phones to learn about price differences in different markets.

2.2 Comparative statics

Objectives and skills

- Understand the distinction between endogenous and exogenous variable, and the meaning of a comparative statics question.
- Use an equilibrium condition for comparative statics analysis.

This section uses a an example to illustrate a comparative statics question, and how it can be answered: "How do transportation costs affect the equilibrium prices and quantities in the two markets?" This question is simple enough to answer without a model: higher transportation costs decrease exports from China to the U.S., increasing supply and decreasing the equilibrium price in China, and having the opposite effect in the U.S. The simplicity enables the reader to focus on the method. Figure 2.1 illustrates the graphical approach to answering this question. The figure shows how moving from zero to positive transport costs shifts down a curve, changing the equilibrium. Mathematics helps for more complicated questions. The first step in addressing a comparative statics question is to be clear about which variables are exogenous (= given, or determined outside the model), and which are endogenous (= determined by the model). Here, the transportation parameter, b, is exogenous, and the prices and quantities in the two markets are endogenous.

The second step is to identify the equilibrium condition that determines the endogenous variables. In this case, the equilibrium condition is the no-arbitrage condition 2.1. The demand functions for the two countries, $p^{\text{china}} = 20 - q^{\text{China}}$ and $p^{\text{U.S.}} = 18 - q^{\text{U.S.}}$, the constraint $q^{\text{U.S.}} = 10 - q^{\text{China}}$, and this equilibrium condition imply

$$20 - q^{\text{China}} = \frac{1}{1+b}\left(18 - \left[10 - q^{\text{China}}\right]\right). \tag{2.2}$$

Once we know q^{China} it is straightforward to find $q^{\text{U.S.}}$ and the two prices, so we consider only the comparative statics of q^{China} with respect to b.

Equation 2.2 gives q^{China} as an *implicit* function of b. Because of its linearity, this equation can be easily solved to yield the *explicit* equation.

$$q^{\text{China}} = \frac{20b + 12}{b+2}.$$

Using the quotient rule, we obtain the derivative:

$$\frac{dq^{\text{China}}}{db} = \frac{28}{(b+2)^2} > 0, \tag{2.3}$$

showing that an increase in transport costs increases equilibrium sales in China. Appendix C shows how to solve more complicated problems, where we are unable to obtain the endogenous variable as an explicit expression of the exogenous parameters.

2.3 Elasticities

Objectives and skills

- Calculate elasticities and understand why it is important that they are "unit free".

Economists frequently use elasticities instead of slopes (derivatives) to express the relation between two variables, e.g. between price and quantity demanded. The slope of the demand function tells us the number of units of change in quantity demanded for a one unit change in price. The elasticity of demand with respect to price tells us the percent change in quantity demanded for a one percent change in price. Unless the meaning is clear from the context, we have to specify the elasticity of *something* (here, quantity demanded) with respect to *something else* (here, price). In general, the value of an elasticity depends on the price (or the "something else") at which it is evaluated.

The symbol dQ denotes the change in Q and the ratio $\frac{dQ}{Q}$ denotes the rate of change; multiplying by 100 converts a rate to a percent, so $\frac{dQ}{Q}100$ is the percent change in Q. The percent change in Q for each one percent change in P is the ratio

$$\frac{\frac{dQ}{Q}100}{\frac{dP}{P}100} = \frac{\frac{dQ}{Q}}{\frac{dp}{p}} = \frac{dQ}{dP}\frac{P}{Q}.$$

The demand function $Q = D(P)$ gives the relation between quantity demanded and price, Q and P. The elasticity of Q with respect to P, denoted η, is defined as

$$\eta = -\frac{dQ}{dP}\frac{P}{Q} = -D'(P)\frac{P}{D(P)}.$$

The convention of including the negative sign in the definition makes the elasticity a positive value. If we want to evaluate this elasticity at a particular price, say P_o, we express it as

$$\eta(P_o) = -D'(P_o)\frac{P_o}{D(P_o)}.$$

Economists frequently use elasticities instead of slopes (derivatives) to describe the response of one variable to a change in another variable, because the elasticity, unlike the slope, is "unit free".[1] The elasticity does not change

[1] Units are important. Columbus underestimated the diameter of the world partly as a consequence of confusion over units. The common belief that Napoleon was quite short (the "little man complex") resulted from confusing English with French units; in fact, he was slightly taller than the average man of his era. The 1999 Mars Climate Orbiter was lost due to miscommunication about units between Lockheed Martin and NASA.

if we change units from pounds to kilos, or measure prices in cents or pesos rather than dollars. The derivative does change if we change units.

To illustrate this difference, suppose that if quantitate is measured in pounds and price in dollars, the demand and inverse demand functions are:

$$\text{demand: } Q = 3 - 5P \quad \text{inverse demand: } P = \frac{3 - Q}{5}.$$

The elasticity of demand with respect to price, evaluated at $P = 0.5$ is

$$-\frac{dQ}{dP}\frac{P}{Q} = 5\frac{0.5}{3 - 5(0.5)} = 5. \tag{2.4}$$

If we measure price in cents ($\tilde{P}$) per pound instead of dollars per pound (P), then using $\tilde{P} = 100P$, the demand and inverse demand functions are

$$\text{demand: } Q = 3 - 0.05\tilde{P} \quad \text{inverse demand } \tilde{P} = 60.0 - \frac{1}{0.05}Q.$$

Changing units from dollars to cents changes both the vertical intercept and the slope of the inverse demand function, altering demand function's appearance, but not the information it contains. Calculating the elasticity of demand at $\tilde{P} = 0.5 \times 100 = 50.0$ returns the same value, 5, as in equation 2.4. Changing the units of a variable changes the derivative, but not the elasticity.

The magnitude of the elasticity of demand depends on characteristics of the good, income, and prices of complements and substitutes. The elasticity of demand for necessities, such as food staples, tends to be small: except in extreme circumstances, a 10% increase in the price of rice causes only a small drop in the demand for rice. In contrast, the elasticity of demand for luxuries may be large.

2.4 Competition and monopoly

Objectives and skills

- Write the payoffs and the equilibrium conditions for a competitive industry and a monopoly.
- Understand the similarities and the differences between these two market structures and their equilibrium conditions.

- Understand the meaning of "marginal revenue", and write it using the elasticity of demand.

The competitive firm takes the market price as given; the monopoly understands that price responds to sales. To understand the effect of moving from competition to monopoly, it is important to hold everything else constant: the inverse demand function, $p(Q)$, and the industry cost function, $c(Q)$, are the same in the two types of markets, where Q is aggregate sales. We compare outcomes under the competitive and the monopoly by comparing the necessary conditions to their profit maximization problems.

The industry and firm cost functions What does it mean to say that an industry, consisting of many firms, has a particular cost function? The simplest way to think of this is to imagine that the industry consists of a large number, n, of factories. Under monopoly, a single firm owns all factories; under the competitive structure, each firm own a single factory. Suppose that the cost of producing q in a single factory is $\tilde{c}(q)$. All firms in the competitive industry are identical, making it is reasonable to assume that they all produce the same quantity, $\frac{1}{n}$'th of industry quantity, so $nq = Q$. If each factory produces $q = \frac{Q}{n}$, then the cost in each firm is $\tilde{c}(q) = \tilde{c}\left(\frac{Q}{n}\right)$. The total industry cost is then $n\tilde{c}\left(\frac{Q}{n}\right)$. We can *define* the industry cost as

$$c(Q) \equiv n\tilde{c}\left(\frac{Q}{n}\right).$$

(The symbol "$\equiv$" means "is defined as".) Taking the derivative of both sides of this equation, with respect to Q, using the chain rule, gives

$$c'(Q) \equiv \frac{dc}{dQ} = n\frac{d\tilde{c}}{d\frac{Q}{n}}\frac{d\frac{Q}{n}}{dQ} = n\frac{d\tilde{c}}{dq}\frac{1}{n} = \frac{d\tilde{c}}{dq} \equiv \tilde{c}'(q) \tag{2.5}$$

Equation 2.5 states that the marginal cost of the *industry* (the expression on the left) equals the marginal cost of the *firm* (or factory), the last expression. This relation holds for any number of firms, n, provided that $q = \frac{Q}{n}$. The cost and marginal cost of producing an arbitrary amount does not change with the market structure. A change in the market structure alters the equilibrium amount produced, but not the technology and therefore not the relation between costs and output.

The competitive equilibrium A representative firm chooses output, q, to maximize profits, $pq-\tilde{c}(q)$, taking price as given. The first order condition is:

$$\frac{d(pq-\tilde{c}(q))}{dq}=p-\tilde{c}'(q)\overset{\text{set}}{=}0\Rightarrow p=\tilde{c}'(q).$$

Using the relation in equation 2.5, we replace $\tilde{c}'(q)$ with $c'(Q)$; recognizing that the price depends on aggregate sales, we replace p with the inverse demand function $p(Q)$ and then rewrite the optimality condition for the representative firm as

$$\text{price} = \text{industry marginal cost: } p(Q)=c'(Q). \tag{2.6}$$

The monopoly equilibrium The monopoly recognizes that the price, $p(Q)$, depends on its sales. It chooses output, Q, to maximize profits, $p(Q)Q-c(Q)$, yielding the first order condition

$$\text{Marginal revenue} = \text{Marginal cost: } p(Q)\left(1-\frac{1}{\eta(Q)}\right)=c'(Q). \tag{2.7}$$

Box 2.1: Derivation of equation 2.7. The first order condition for profit maximization is

$$\begin{aligned}\tfrac{d(p(Q)Q-c(Q))}{dQ}&=p'(Q)Q+p-c'(Q)\overset{\text{set}}{=}0\\ &\Rightarrow p'(Q)Q+p=c'(Q),\end{aligned}$$

which states that marginal revenue, $p'(Q)Q+p$, equals marginal cost, $c'(Q)$. We can write marginal revenue (denoted $MR(Q)$) using the elasticity of demand, η:

$$\begin{aligned}MR(Q)\equiv\tfrac{d[p(Q)Q]}{dQ}&=p'(Q)Q+p=p\left(1+\tfrac{dp}{dQ}\tfrac{Q}{p}\right)\\ p\left(1+\tfrac{1}{\frac{dQ}{dp}\frac{p}{Q}}\right)&=p(Q)\left(1-\tfrac{1}{\eta(Q)}\right).\end{aligned}$$

Comparing the two equilibria The equilibrium conditions for the competitive industry and the monopoly are

$$\begin{aligned}\text{Competition: }& p(Q)=c'(Q)\\ \text{Monopoly }& p(Q)\left(1-\tfrac{1}{\eta(Q)}\right)=c'(Q).\end{aligned}$$

Both of these equations involve the inverse demand function, $p(Q)$, and the industry marginal cost function, $c'(Q)$. The competitive equilibrium condition sets price equal to the industry marginal cost, and the monopoly condition sets marginal revenue equal to industry marginal cost. Once we have the necessary condition for a competitive equilibrium, we can (in many cases) obtain the necessary condition for a monopoly equilibrium *merely by replacing p with $p\left(1-\frac{1}{\eta}\right)$.*[2]

The more elastic is the demand function that the monopoly faces (the larger is η), the less opportunity the monopoly has to exercise market power: it understands that any effort to raise the market price requires a large reduction in sales, and a corresponding fall in revenue. As the market demand becomes infinitely elastic, i.e. as $\eta \to \infty$ ("η goes to infinity"), the monopoly looses all market power, and behaves like a competitive firm. The monopoly never produces where $\eta < 1$; at such a point, marginal revenue is negative. Examples $1-3$ at the end of this chapter show how to set up the objective functions for the competitive and monopoly industries, use the first order conditions to the two maximization problems to find equilibrium price and quantity in the two cases, and then examine the effect of the elasticity of demand on difference in equilibrium price.

Optimization and equilibrium In most economic contexts, an equilibrium occurs where all agents simultaneously solve their optimization problems: no one wants to move *unilaterally* away from an equilibrium. At a competitive equilibrium consumers maximize utility, resulting in quantity demanded on the demand function, and producers maximize profits, resulting in quantity supplied on the supply function. Markets clear, so supply equals demand.

Models help to clarify complex situations, but do not literally describe behavior. If people behave completely randomly, then optimization-based economic models are useless. If people attempt to pursue their self-interest, and behave with a modicum of rationality, these models are informative.

[2] Early in the study of arithmetic we learn that the order in which operations are performed matters: $(3+4)\times 7 = 49 \neq 3 + (4 \times 7) = 31$. The order of operations also matters in carrying out economic calculations. For the monopoly, we *first* replace "price" with the inverse demand function, and *then* we take the derivative of profit with respect to sales. For the competitive firm, we *first* take the derivative of profit with respect to sales, taking price as given, and *then* substitute the inverse demand function into the first order condition to obtain an equation in quantity. *The order of these two steps is critical.*

Firms that are consistently irrational are not likely to survive long in the market place. Shoppers may not buy food that is good for them, but they buy the food that they want, and in that respect they act in their self-interest.

Equilibrium is also an abstraction. Competitive firms' optimal production level depends on the price they expect to receive, but their expectations may be wrong. If they have already committed a certain quantity to the market, but the price is lower than they expected, the price-quantity point is below the supply curve. Markets are unlikely to be exactly in equilibrium, but people respond to their mistakes, and those responses likely move a market towards equilibrium. If firms find that the price has repeatedly been lower than their marginal cost, they have an incentive to decrease quantity, causing price to rise and the outcome to move toward equilibrium.

Optimality and no-arbitrage conditions We used only basic economic logic to obtain the no arbitrage condition 2.1. We can also obtain this equation as the first order condition to an optimization problem. Using the constraint $q^{\text{U.S.}} = 10 - q^{\text{China}}$, profits for the price-taking tea seller (revenue in China plus revenue in the US minus transportation costs) equal

$$\pi = p^{\text{China}} q^{\text{China}} + p^{\text{U.S}} \left(10 - q^{\text{China}}\right) - bp^{\text{China}} \left(10 - q^{\text{China}}\right).$$

The first order condition (at an interior equilibrium) is

$$\frac{d\pi}{dq^{\text{China}}} = p^{\text{China}} - p^{\text{U.S}} + bp^{\text{China}} \overset{\text{set}}{=} 0.$$

Rearranging the last equation produces the no-arbitrage condition 2.1.

2.5 Discounting

Objectives and skills

- Understand the rationale for and the implementation of discounting.
- Use discount factors to calculate the present value of a "stream" (= sequence) of future costs or payments.
- Understand the relation between the magnitude of a discount rate and the length of a period over which discounting occurs.

After explaining discounting, we provide three examples of its use. A box of tea in China and a box of tea in the U.S. are not the same commodity if it is costly to move the box from one location to the other. Similarly, a dollar ten years from now is not the same as a dollar today, because there is an opportunity cost, the foregone investment opportunity, of receiving the dollar later rather than earlier. Resource management may involve deciding when to take a unit of oil out of the ground, or when to harvest a unit of fish, requiring the manager to compare the value to the firm of extracting at different points in time. This comparison involves discounting.

Pick a period of arbitrary length, say one year. Suppose that the most profitable riskless investment available pays a positive return of r after one year. We call r the discount rate. If a person at period 0 invests z dollars for one year in an asset that returns the rate r, then at the end of the year the person has $z(1+r)$. The person with this investment opportunity is indifferent between receiving \$1 at the beginning of the next year and \$$z$ in the current period, if and only if $z(1+r) = 1$, i.e. if and only if $z = \frac{1}{1+r}$. We define $\rho = \frac{1}{1+r}$ as the discount factor. This person is indifferent between receiving \$43.60 at the beginning of next period, or \$43.60 $\times \rho$ at the beginning of the current period. Multiplying an amount received one year in the future by ρ, produces the "present value" of the future receipt.

A person is indifferent between receiving one dollar at the end of two years and \$$z$ today if and only if $(1+r)(1+r)z = 1$, i.e. if and only if $z = \rho^2$. Thus, ρ^2 is the present value today of a dollar in two years. Similarly, ρ^n is the present value today of a dollar n years from now.[3]

If a firm obtains profits π_t during periods $t = 0, 1, 2...T$, then the present discounted stream of profits equals $\sum_{t=0}^{T} \rho^t \pi_t$. In the special case where $\pi_t = \pi$, a constant, we can simplify this sum using the formula for a geometric series:

$$\sum_{t=0}^{\infty} \pi \rho^t = \pi \sum_{t=0}^{\infty} \rho^t = \frac{\pi}{1-\rho}. \tag{2.8}$$

[3] The number of periods, n, it takes to double the value of an investment depends on the discount (= interest) rate, using the formula $(1+r)^n = 2$, or (taking logs and simplifying) $n = \frac{\ln 2}{\ln(1+r)}$. For $r = 0.01$ (a 1% interest rate), $n \approx 70$; for $r = 0.1$, $n \approx 7$.

(π represents profits, not the number $3.14...$.) This formula implies

$$\sum_{t=0}^{T} \pi\rho^t = \sum_{t=0}^{\infty} \pi\rho^t - \sum_{t=T+1}^{\infty} \pi\rho^t = \sum_{t=0}^{\infty} \pi\rho^t - \rho^{T+1} \sum_{t=0}^{\infty} \pi\rho^t = \frac{1-\rho^{T+1}}{1-\rho}\pi. \tag{2.9}$$

Relation between the discount rate and the length of a period The numerical value of the discount rate, and thus of the discount factor, depends on the length of a period of time. If a period lasts for ten years, and an asset held for one ten-year period pays a return of $\tilde{r}$, then one dollar invested in this asset for three periods (30 years) returns $(1+\tilde{r})^3$. We say that this return is "compounded" every decade. We can convert this decadal return to an annual return by choosing the annual discount rate, r to satisfy $(1+r)^{10} = (1+\tilde{r})$. Taking logs of both sides and simplifying gives $\ln(1+r) = \frac{\ln(1+\tilde{r})}{10}$. If $\tilde{r} = 0.8$, then $r = 0.061$; this asset pays an 80% return over a decade, or a 6.1% return compounded annually. (We multiply by 100 to translate a rate into a percentage.) Given the decadal and annual discount rates, the corresponding decadal and annual discount factors are $\tilde{\rho} = \frac{1}{1+\tilde{r}}$ and $\rho = \frac{1}{1+r}$.[4] The following three examples illustrate the use of discounting.

The levelized cost of electricity Electricity can be produced using different production methods and different inputs, leading to different cost streams and producing different amounts of power. Nuclear-powered plants are expensive to build and require decommissioning, but have low fuel costs. Fossil fuel plants have relatively low investment and decommissioning costs, but high fuel costs. The "levelized cost" of electricity (LCOE) provides a basis for comparing the cost of producing electricity using different methods. The data for this calculation includes estimates of the year-t capital cost, C_t, variable cost, V_t (including fuel and maintenance), the amount of energy produced, E_t, and the lifetime (including construction and decommissioning

[4] Under *continuous discounting* at rate $\tilde{r}$, the discount factor after t units of time is $e^{-\tilde{r}t}$. If a unit of time equals one year, and the annual discount rate is r, then the continuous rate $\tilde{r}$ satisfies $e^{-\tilde{r}} = \frac{1}{1+r}$. Taking logs of both sides gives $\tilde{r} = \ln(1+r)$. If $r = 0.05$ (5% annual discount rate), $\tilde{r} = \ln(1.05) = 0.049$.

time) of the project, $n+1$ years. The formula for LCOE is

$$LCOE = \frac{\sum_{t=0}^{n} \rho^t \left(C_t + V_t\right)}{\sum_{t=0}^{n} \rho^t E_t}. \tag{2.10}$$

The LCOE often excludes potentially important considerations: wind or solar may require significant network upgrades in order to bring the power to market; fossil fuels have health and climate-related externalities; nuclear power creates the risk of rare but catastrophic events. Incorporating these and other considerations requires additional data.

The example in Table 2.1 illustrates the calculation and shows the sensitivity of LCOEs to the discount rate. Type A generation method is expensive to construct but cheap to run and lasts a long time. Type B is cheap to build, expensive to run, and has a shorter lifetime. They both produce the same amount of energy per year (one unit).

	capital cost (billion \$)	annual operating cost	lifetime	construction time
Type A	2	0.05	45 years	5 years
Type B	0.3	0.12	30 years	2 years

Table 2.2: Example: two different types of power plants

For this example, the LCOE of the two power plants are

$$LCOE^A = \frac{\left(2 + \sum_{t=5}^{50} \rho^t \left(0.05\right)\right)}{\sum_{t=5}^{50} \rho^t 1} \text{ and } LCOE^B = \frac{\left(0.3 + \sum_{t=2}^{32} \rho^t \left(0.12\right)\right)}{\sum_{t=2}^{32} \rho^t 1}$$

Figure 2.2 shows the ratio of the two levelized costs as a function of the discount rate. The costs are equal (the ratio is 1) for $r = 2.7\%$; Type A is 10% cheaper at $r = 2\%$ and 17% more expensive at $r = 4\%$. When "money is cheap" (the interest rate is low), it is economical to use the method that has large up-front costs but lower costs overall (Type A). However, when the interest rate is high, it is economical to use Type B, which has lower initial costs but higher undiscounted total costs.

Table 2.2 shows U.S. estimates for several power sources.

Conventional Coal= 96	IGCC* =116	Natural Gas CCC** = 66	Advanced Nuclear = 96
Wind = 80	Wind offshore = 204	Solar PV2 = 130	Hydro = 85

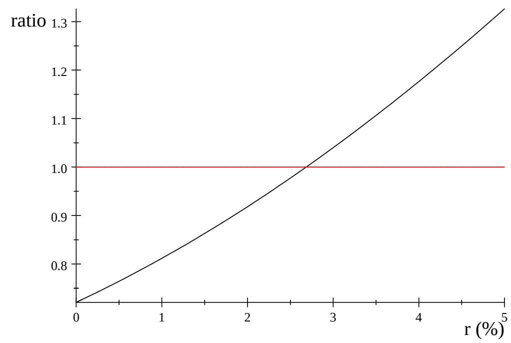

Figure 2.2: Ratio of LCOE $\frac{\text{Type A}}{\text{Type B}}$.

Table 2.2 Estimated LCOE (2012$/MWh) for plants entering service in 2019. *Integrated Coal-Gasification Combined Cycle. ** Conventional Combined Cycle (U.S. Energy Information Administration, 2014)

The social cost of carbon The "social cost of carbon" (SCC) is defined as the present discounted stream of damages due to a unit of carbon emissions today. It plays an important role in climate economics (Chapter 19); the US Environmental Protection Agency (EPA) uses the SCC in cost/benefit analyses of rules and legislation affecting greenhouse gas emissions. The SCC depends on: the relation between a unit of emissions today and future carbon stocks; the relation between carbon stocks and temperature changes; the relation between temperature changes and economic damages; and the discount rate. Higher discount rates (lower discount factors) place less weight on future damages, and therefore lead to a lower SCC. With a discount rate of 2.5%, the EPA estimates the SCC in 2015 at $11 per metric ton of CO_2, rising to $56 for a 2.5% discount rate: halving the discount rate increases the SCC by a factor of five.

These estimates involve complex models, but an example shows the role of discounting. Suppose that each metric ton of atmospheric CO_2 creates d dollars of annual economic damage, and that carbon decays at a constant rate δ.[5] With these assumptions, one unit of emissions today increases the carbon stock t periods from now by $(1-\delta)^t$ and creates $d\,(1-\delta)^t$ dollars of

[5] Carbon does not literally decay. CO_2 is emitted to the atmosphere, and over time some of it moves to different oceanic and terrestrial "reservoirs". The model of constant decay is one of the simplest ways to approximate carbon leaving the atmosphere.

damage in period t. The present discounted value of the stream of damages due to this unit of emissions equals

$$SCC = \sum_{t=0}^{\infty} d\left(1-\delta\right)^{t} \rho^{t} = \frac{d}{1-\left(1-\delta\right)\rho} = \frac{d}{r+\delta}\left(r+1\right).$$

The second equality uses formula 2.8 and the third uses the definition of ρ. The smaller is the discount rate (the larger is ρ), the larger is the SCC.

Implicit subsidies from ignoring discounting The US Reclamation Act of 1902 used receipts from the sale of federal lands to finance the Reclamation Fund, which paid for irrigation projects in western states. The Fund was designed to be self-perpetuating, with the settlers who used the water repaying the cost of the project, without interest. The settlers were thus given a no-interest loan. The repayment period was initially ten years, but later projects were financed over 40 - 50 years. The implicit subsidy arising from these no-interest loans could be as high as 90% of the cost of the project, depending on the length of the repayment period and the interest rate (the opportunity cost of money).

If users repay, without interest, the cost of a project, C, over a period of T years; their annual repayment is $\frac{C}{T}$. If the opportunity cost of funds is r, the value of this stream of payments is (using equation 2.9) $\frac{C}{T}\frac{1-\rho^{T}}{1-\rho}$. The subsidy, as a percent of the cost of the project, C, is

$$S = \frac{C - \frac{C}{T}\frac{1-\rho^{T}}{1-\rho}}{C}100 = \left(1 - \frac{1-\rho^{T}}{T\left(1-\rho\right)}\right)100.$$

Figure 2.3 shows that the subsidy is sensitive to both the repayment period, T, and the interest rate. Low- or no-interest loans can result in large subsidies.

2.6 Welfare

Objectives and skills

- Understand the meaning of "Pareto efficient".
- Be familiar with the two Fundamental Welfare Theorems.

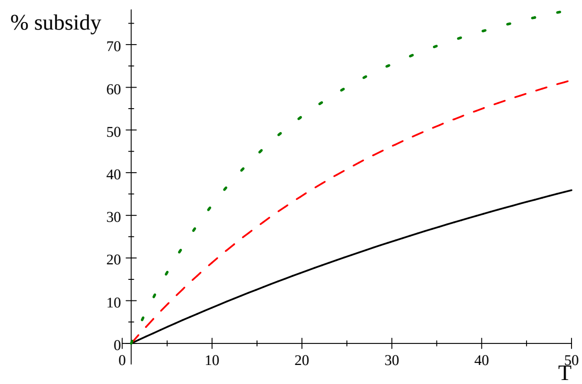

Figure 2.3: Percent subsidy as a function of repayment time, T years, for $r = 0.02$ (solid) and $r = 0.04$ (dashed), and $r = 0.1$ (dashed)

Resource economics studies the allocation of a natural resource over time. Under certain conditions, the allocation under competitive markets is Pareto efficient, meaning that there does not exist another allocation of the resource that makes at least one agent better off, without making any agent worse off. "Pareto efficient" is not a value judgement; "efficient" does not mean "good". If Jiangfeng and Mary get utility only from their own consumption of a good with fixed supply equal to one unit, a *feasible* allocation gives Jiangfeng $z \geq 0$ and Mary $w \geq 0$ units of the good, with $z + w \leq 1$. An *efficient* allocation makes sure that all of the good is consumed; *any* outcome with $z + w = 1$ is efficient. Efficiency is a rather weak criterion. We might prefer the equal but inefficient allocation, $z = w = 0.4999$ to the efficient but ethically questionable allocation $w = 1$, $z = 0$.

Chapters 4 – 5 emphasize the competitive equilibrium without market failures. Firms own natural resource stocks and choose how much to extract and sell in each period. In order to make statements about the social welfare in a competitive equilibrium, we must decide how to measures welfare. Given a welfare criterion, we chooses how much to extract and sell in each period in order to maximize welfare. We can then compare the outcome under competition (or monopoly) with the outcome under this social planner.

We use partial equilibrium models: those that involve a single market, e.g. the market for a particular resource. These models take as given all "outside" considerations that influence this market. If the resource is petroleum, the partial equilibrium model seeks to explain petroleum prices and quantities, over time, taking as given: levels of income (which affect demand for petroleum); prices of factors used to produce petroleum (labor, machinery); prices of substitutes (natural gas) and complements (cars); and technology (drilling techniques). With a partial equilibrium model, consumer and producer surplus are reasonable measures of consumer and producer welfare, and their sum is a reasonable measure of social welfare in a period. We take the social welfare function to be the discounted sum (over time) of welfare in each period. This criterion, known as "discounted utilitarianism", is used in most resource models. Our fictitious social planner is a discounted utilitarian.

Dynamics and market failures are both important for many natural resources. We begin by studying dynamics when there are no market failures (except possibly monopoly). We then discuss market failures, ignoring dynamics. With these building blocks, we turn to the case of interest, where there are both dynamics and market failures. We say that markets are "complete" if there is a market for every type of transaction that people would like to make. For example, if someone would like to buy water and someone else is willing to sell water, then "complete markets" requires that there is actually a water market that makes their exchange possible (Chapter 17). An "unpriced externality" is a consequence of the market transaction not fully reflected in the price of the good. For example, the price of fossil fuels does not include the environmental damage (the unpriced externality) arising from extracting and using the fuels. The following result is the starting point for welfare economics:

> *First Fundamental Welfare Theorem*: If markets are complete, there are no unpriced externalities, and agents are price-takers, then any competitive equilibrium is Pareto efficient.

A second theorem provides conditions under which a particular Pareto efficient outcome can result from a competitive equilibrium. We say that a set of transfers and taxes "supports" a particular "outcome X" in a competitive equilibrium if, in the presence of those transfers and taxes, the competitive equilibrium has the same prices and quantities as "outcome X". The second theorem is

> *Second Fundamental Welfare Theorem*: Provided that a technical requirement ("convexity") is satisfied, any Pareto efficient.[6]

[6]Constant and decreasing returns to scale technologies are "convex" and increasing

These two theorems form the basis for understanding the welfare properties of a competitive equilibrium. The first theorem gives conditions under which the competitive equilibrium is efficient. The second gives conditions under which we can obtain any other efficient outcome, as a competitive equilibrium, by using appropriate taxes and transfers.

2.7 Summary

An example of arbitrage over space illustrates the meaning of arbitrage. Many of the main ideas in this book are based on arbitrage over time: instead of selling the commodity in one country rather than another, the firm sells it at one point in time rather than another. Understanding spatial arbitrage makes it easy to understand intertemporal arbitrage.

Economic models help to determine how a change in an exogenous parameter changes an endogenous variable. This kind of question is known as a comparative statics question. Casual reasoning or graphical methods suffice to answer easy comparative statics questions. In more complicated cases, we use mathematics, beginning with an equilibrium condition (e.g., supply equals demand). One approach uses this condition to find an explicit expression for the endogenous variable, as a function of the exogenous variables. An alternative uses the differential of the equilibrium condition.

We defined elasticities, and illustrated the definition using the elasticity of demand with respect to price. It is important to be able to calculate an elasticity, and to understand why it is unit free.

In order to compare perfect competition and monopoly, we want to "hold everything else constant", apart from the market structure. In this context, we require that the demand and cost *functions* (not their *levels*) are the same for both market structures. We can think of the industry consisting of many factories. In the competitive environment, each firm owns one of these factories, and in the monopoly, a single firm owns all factories.

Both the monopoly and the representative firm want to maximize profits. For the price-taking competitive firms, the equilibrium condition is "price equals marginal cost". The monopoly understands that its sales affect the price; the monopoly marginal revenue equals $p\,(1-1/\eta)$, where η is the elas-

returns to scale technologies are not. Decreasing/constant/increasing returns to scale mean that doubling all inputs: less than doubles/ exactly doubles/ more than doubles output.

ticity of demand. Provided that η is finite, the monopoly sells less than the competitive industry. As demand becomes more elastic, the monopoly has less market power.

We use the discount factor to compare money (e.g. profits) received in different periods. The discount factor is $\rho = 1/(1+r)$, where r is the discount rate, equal to the highest riskless return available to the agent. The discount factor converts future values into present values.

An outcome, such as the allocation of a product across individuals, geographical regions, or time, is Pareto efficient if there is no reallocation that makes some agent better off without making any agent worse off. The two fundamental welfare theorems describe the relation between a competitive equilibrium and the outcome under a social planner. The first of these theorems provides conditions under which a competitive equilibrium is Pareto efficient. The second provides conditions under which any Pareto efficient equilibrium can be supported as a competitive equilibrium by means of taxes or income transfers.

2.8 Terms, examples, study questions, and exercises

Terms and concepts

Demand function, inverse demand function, arbitrage, no-arbitrage condition, interior equilibrium, iceberg transportation costs, endogenous and exogenous variable, law of one price, implicit function, explicit function, comparative statics, differential, first order condition, marginal revenue, order of operations, discount function, discount factor, opportunity cost, compounded, capital cost, operating cost, decommissioning cost, levelized cost, partial equilibrium, externality, complete markets, efficient, Pareto efficient, consumer and producer surplus, feasible, discounted utilitarianism, taxes and transfers "supporting an outcome".

Examples

These examples use a "constant elasticity of demand function" $Q = \left(\frac{p}{A}\right)^{-\eta}$, with the inverse demand $p(Q) = AQ^{-\frac{1}{\eta}}$; here, the elasticity η is a constant (a parameter), not a function of price.

Example 1 *This example shows how to write the maximization problem for the competitive industry and solve for the equilibrium. Inverse demand is* $p(Q) = AQ^{-\frac{1}{\eta}}$ *and the industry cost is* $c(Q) = \frac{b}{2}Q^2$. *The objective, first order condition, and industry equilibrium condition of the competitive industry are*

$$\begin{aligned} \textit{objective: max} & \quad {}_Q \left[pQ - \frac{b}{2}Q^2\right] \\ \textit{first order condition} \quad : & \quad p - bQ \overset{set}{=} 0 \qquad (2.11) \\ \textit{equilibrium condition} \quad : & \quad AQ^{-\frac{1}{\eta}} - bQ \overset{set}{=} 0 \Rightarrow \qquad (2.12) \\ Q &= \left(\frac{A}{b}\right)^{\frac{\eta}{1+\eta}} \Rightarrow p = A\left(\frac{A}{b}\right)^{\frac{-1}{1+\eta}}. \end{aligned}$$

The first order condition states that price equals marginal cost. We obtain the equilibrium condition by replacing price with the inverse demand function. The last line solves the equilibrium condition for both equilibrium quantity and equilibrium price.

Example 2 *This example uses the same inverse demand function and cost function to describe the monopoly equilibrium. Here we assume that* $\eta > 1$, *so that marginal revenue is positive. The objective and first order condition for the monopoly is*

$$\begin{aligned} \textit{objective} \quad : & \quad \max_Q \left[AQ^{1-\frac{1}{\eta}} - \frac{b}{2}Q^2\right] \\ \textit{first order/equilibrium condition} \quad : & \\ \left(1 - \frac{1}{\eta}\right)AQ^{-\frac{1}{\eta}} - bQ &= \left(1 - \frac{1}{\eta}\right)p - bQ \overset{set}{=} 0 \qquad (2.13) \end{aligned}$$

Note that we can obtain equation 2.13 by replacing "price" in 2.12 with "marginal revenue" which equals $\left(1 - \frac{1}{\eta}\right)p$. *Solving equation 2.13 gives*

$$\begin{aligned} \left(1 - \frac{1}{\eta}\right)AQ^{-\frac{1}{\eta}} &= bQ \Rightarrow Q = \left[\frac{A\left(1 - \frac{1}{\eta}\right)}{b}\right]^{\frac{\eta}{\eta+1}} \Rightarrow \\ p &= A\left[\frac{A\left(1 - \frac{1}{\eta}\right)}{b}\right]^{\frac{-1}{\eta+1}} \end{aligned}$$

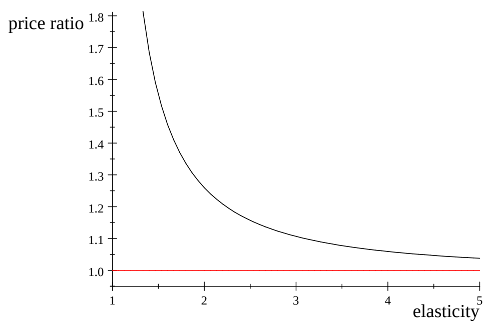

Figure 2.4: Ratio of monopoly to competitive price

Example 3 *This example shows how the demand elasticity, η, affects the monopoly's ability to exercise market power, as measured by the ratio of the equilibrium monopoly to competitive price. The ratio of monopoly to competitive price for this example is*

$$ratio = \frac{A\left(\frac{A\left(1-\frac{1}{\eta}\right)}{b}\right)^{\frac{-1}{\eta+1}}}{A\left(\frac{A}{b}\right)^{\frac{-1}{1+\eta}}} = \frac{\left(\frac{1}{b}\right)^{\frac{1}{\eta+1}}}{\left(\frac{1}{b\eta}(\eta-1)\right)^{\frac{1}{\eta+1}}} = \eta^{\frac{1}{\eta+1}}(\eta-1)^{-\frac{1}{\eta+1}}$$

Figure 2.4 graphs the ratio (which is independent of the parameters A and b). For low elasticity of demand (η close to 1) monopoly power is substantial, and the monopoly price is much larger than the competitive price. (The ratio of prices is large.) For large elasticity of demand, the monopoly has little market power, and the monopoly and competitive prices are similar.

Study questions

1. You should be able to use the type of figure in Section 2.1 to illustrate the effect of a change on demand in one country, or a change in available supply, or in transportation costs, on the equilibrium allocation of sales across country.

2. Given inverse demand functions in the two countries, available supply, and the transport costs, you should be able to write down the equilibrium condition and to write a comparative statics expression showing

the effect of a change in an exogenous variable on an endogenous variable. You should be clear about the distinction between exogenous and endogenous variables.

3. You should know what an elasticity is, how to calculate it, and what it means to say that the elasticity is unit free.

4. Given an industry cost function and an inverse demand, you should be able to write down the equilibrium conditions that determine sales under competition and under monopoly.

5. You should know the relation between a discount rate and a discount factor, and understand what they are used for. Given the formula for the sum of a geometric series, you should be able to calculate the present discounted stream of payoffs. You should be able to work through an example like the first one in the text that compares the cost of two methods of electricity generation.

6. You should understand the meaning and be able to describe the two Fundamental Welfare theorems.

Exercises

1. When quantities are measured in pounds and prices in dollars, the demand function is $Q = 3 - 5P$. (a) What is the elasticity of demand, evaluated at $P = 0.5$? (b) Express the same relation between demand and price, using different units, q and p: q are in units of kilos, and p in pesos. (Which equation is correct, $q = 2.2Q$ of $q = \frac{Q}{2.2}$?) There are 3 pesos per dollar. (Which equation is correct, $p = 3P$ or $p = \frac{P}{3}$?) (c) Using the new units, q and p, express the elasticity of demand with respect to price, evaluated at the price of one dollar. (d) What is the point of this exercise?

2. How does an increase in transportation costs affect the location of the dotted line in Figure 2.1, and how does this change alter the equilibrium price and the allocation of tea between the two countries?

3. How does an increase in the available supply (e.g. from 10 units to 12 units of tea) change the appearance of Figure 2.1, and how does this

change in supply alter the equilibrium quantities and prices in the two countries?

4. Calculate $\eta(Q)$ for the demand function $Q = a - hp$ and then for the demand function $Q = ap^{-h}$, where h is a positive number. Graph the two elasticities as a function of Q.

5. Suppose that a monopoly owns the ten units of tea in China. There is no transportation costs ($b = 0$). Using the inverse demand functions in Section 2.1, find monopoly sales two markets. How does the monopoly sales in China compare to sales by a competitive firm?

6. Consider the monopoly in the previous question. Suppose that ice-berg transportation costs are b. Using the equilibrium condition for the monopoly, find the comparative statics of the monopoly's sales in China, with respect to b. (Write down the equilibrium condition for the monopoly, solve for sales in China, and take the derivative of this expression with respect to b.)

7. Suppose that the industry has the cost function $c(Q) = 2Q + \frac{3}{2}Q^2$. This industry consists of n firms, each with cost function $c(q)$. Find $c(q)$. Hint: "Guess" that the single firm's cost function is of the form $\tilde{c}(q) = aq + \frac{b}{2}q^2$. Then use the requirement that $c(Q) = n\tilde{c}\left(\frac{Q}{n}\right)$, to write
$$2Q + \frac{3}{2}Q^2 = n\left(a\frac{Q}{n} + \frac{b}{2}\left(\frac{Q}{n}\right)^2\right).$$
This relation must hold for all Q (not just a particular Q), so we can "equate coefficients" of Q and Q^2 to find the values of a and b.

8. A plant that supplies 1 unit of electricity per year, costs \$1 billion to build, lasts 25 years, and has an annual operating cost of \$0.2 billion; it costs \$0.1 billion to decommission the plant at the end of its lifetime (25 years). (Assume that the construction costs and the operating costs are paid at the beginning of the period, and that the decommissioning cost is paid at the end of the life of the plant.) The annual discount rate is r, with discount factor $\rho = \frac{1}{1+r}$. Write the formula for the present value of the cost of providing 1 unit of electricity for 100 years, including the decommissioning costs. (Hint: First find the present value of providing one unit of electricity for 25 years. Denote this

magnitude as Z. Then find the present value of incurring this cost, Z, 4 times: in periods 0, 25, 50, and 75.)

9. A monopoly faces the demand function $q = p^{-1.2}$ and has production costs $c(q) = \frac{b}{2}q^2$. Find the comparative statics of equilibrium sales, with respect to the cost parameter, b. (Hint: Write the monopoly's first order condition (marginal revenue equals marginal cost). Solve this expression for q as a function of b, and take the derivative of q with respect to b.)

10. A person plans to save \$1 for 20 years. They can invest at an annual rate of 10% ($r = 0.1$). This investment opportunity "compounds annually" (meaning that they receive interest payments at the end of each year). A second investment opportunity pays a return of $\tilde{r} \times 100\%$, compounded every decade. (After one decade, the investment of one dollar yields $1 + \tilde{r}$.) For what value of $\tilde{r}$ is the person indifferent between these two investments? (Assume that there is no chance that the person wants to cash in the investment before the 20 year period.) Explain the rationale behind your calculation.

11. (*) This exercise illustrates the First Fundamental welfare theorem. Inverse demand equals $p(q)$ and total cost is $\frac{1}{2}cq^2$. (a) Write the condition for equilibrium in a competitive market. (b) For a linear demand function, draw the graphs whose intersection determines the competitive equilibrium. Using this graph, identify consumer and producer surplus. (c) Explain in words why consumer surplus equals
$$\int_0^q p(w)\,dw - p(q)\,q.$$
(d) Write the expression for producer surplus, equal to revenue minus costs. (e) Define social surplus, $S(q)$, as the sum of consumer and producer surplus. Write the expression for social surplus. Using Leibniz's rule (see math appendix) write the first order condition for maximizing social surplus, by choice of q. (f) Compare this first order condition with the equilibrium condition under competition. Explain why this comparison implies that the competitive equilibrium and the solution to the social planner's problem (maximizing social surplus) are identical. What does this have to do with the First Fundamental Welfare Theorem?

Sources

The U.S. Energy Information Administration (2014) presents and explains estimates of the Levelized Cost of Electricity.

The Interagency Working Group (2013) presents estimates of the Social Cost of Carbon.

Sen *On Ethics and Economics* (1987) New York, Blackwell provides a background on welfare economics.

Aker (2010) provides evidence of the relation between cell phones and agricultural markets.

Glaeser and Kohlhase (2004) provide estimates of transport costs, and discuss their role in international trade.

Prominent estimates of market power in resource markets include: Hnyilicza and Pindyck (1976), Stollery (1985), Pindyck (1987), Ellis and Halvorsen (2001), Cerda (2007).

Chapter 3

Nonrenewable resources

Objectives

- Ability to analyze equilibria under competition and monopoly for the two-period model of nonrenewable resources.

Information and skills

- Translate the techniques and intuition from the "trade in tea" model to the nonrenewable resource setting.
- Understand the relation between transport costs in the trade model and the discount factor in the resource setting.
- Derive and interpret an equilibrium condition and analyze it using graphical methods, for both competition and monopoly.
- Do comparative statics with respect to extraction costs and the discount factor.

A two-period model provides much of the intuition needed to understand equilibrium in a nonrenewable resources market. We use graphical methods to analyze the equilibrium under competition or monopoly when firms are unable to save any resource beyond the second period. We emphasize the case where the initial stock is small enough, relative to demand, that firms want to exhaust the resource during this time.

Arbitrage provides the basis for the intuition in resource models, but here we speak of arbitrage over time, instead of over space. A sales *trajectory* is the sequence of sales, and a price trajectory is the associated sequence of prices. In the two-period setting, each of these sequences contains only two elements. We describe the competitive and the monopoly models, and then explain how to answer the following type of comparative statics question: How does a change in a demand or a cost parameter affect the equilibrium level of first-period sales?

3.1 The competitive equilibrium

Objectives and skills

- Write the objective and the constraints for a competitive firm.
- Obtain and interpret the "no-intertemporal arbitrage" (equilibrium) condition under competition, and analyze it graphically.
- Understand the effect of constant average extraction costs on the equilibrium sales and price trajectories.

Chapter 2.1 considered the allocation of a fixed quantity of tea over two countries, in the presence of iceberg transportation costs. Here, a fixed stock of the resource replaces tea, two periods replace the two countries, and the discount factor replaces the iceberg transportation costs.

We require a bit of notation. The first period is denoted $t = 0$, and the second period, $t = 1$. A price-taking firm has discount factor ρ, faces prices p_0 and p_1 in periods 0 and 1, must pay extraction cost c for each unit extracted, and has a fixed stock of the resource, x units. Apart from the fact that the trade example did not include production costs, the trade and the resource models are the same; we merely call things by different names. The intuition for the equilibrium in the two models is also the same.

Let y be sales in period 0. Assuming that all of the resource is sold, $x - y$ equals period-1 sales. At an interior equilibrium, extraction is positive in both periods: $x > y > 0$. The firm wants to maximize the sum of present value profits in the two periods:

$$\pi^{\text{competitive}}(y; p_0, p_1) = (p_0 - c)\, y + \rho\,(p_1 - c)\,(x - y). \tag{3.1}$$

Multiplying period-1 profits by the discount factor, ρ, gives the present value of period-1 profits. The derivative of $\pi(y; p_0, p_1)$ with respect to y is

$$\frac{d\pi^{\text{competitive}}}{dy} = (p_0 - c) - \rho(p_1 - c).$$

The firm prefers to sell all of its stock in period 0 if $(p_0 - c) > \rho(p_1 - c)$. It prefers to sell all of its stock in period 1 if $(p_0 - c) < \rho(p_1 - c)$. In order for the firm to sell a positive quantity in both periods, as we assume, it must be *indifferent* about when to sell the stock. This indifference requires that the present value of period $t = 1$ price minus marginal cost equals the value of price minus marginal cost in period $t = 0$:

$$\frac{d\pi^{\text{competitive}}}{dy} = 0 \text{ if and only if } (p_0 - c) = \rho(p_1 - c). \tag{3.2}$$

The second equation is a "no-intertemporal arbitrage" condition; it holds in an interior competitive equilibrium. The equation states that the firm cannot increase its profits by moving sales from one period to another.

The competitive resource owner, just like the competitive exporter in the trade example, takes prices as given. These prices adjust in response to the amount of supply brought to market. Actions of an individual resource owner, just like the actions of an individual exporter, have negligible effect on the price. However, all resource owners (in our model) have the same costs and discount factor, so they have the same incentives. Therefore, we can proceed as if there is a "representative firm" that takes price as given, and owns all of the stock in the industry. The price responds to changes in this representative firm's supply.

Figure 3.1 shows the market when extraction costs are $c = 0$ and the inverse demand in both periods is $p = 20 - y$. Sales in period 0 equal y. With initial stock $x = 10$, period-1 sales equal $10 - y$. The solid line shows the demand function in period 0: as y increases, the equilibrium price falls. The dashed line shows the demand function in period 1: as y increases, period 1 sales, $10 - y$, fall, so the price in that period rises. If the discount rate is 0 $(r = 0)$ then the discount factor, $\rho = \frac{1}{1+r}$, equals 1. In this scenario, firms allocate the stock evenly between the two periods, and the price in each period equals 15. With zero discount rate, the firm is indifferent between selling in periods 0 or 1 if and only if the prices in the two periods are equal.

The dotted line shows the present value (in period 0) of the period-1 price if the discount rate is $r = 0.3$, so $\rho = \frac{1}{1.3} = 0.77$. The equilibrium occurs

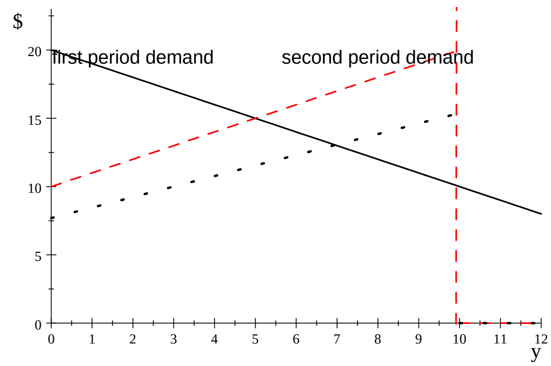

Figure 3.1: Demand in period 0 (solid). Demand in period 1 (dashed). Present value of period 1 price (dotted).

where the *present value* of price is the same in both periods, i.e. where the solid and the dotted lines intersect. With our demand function and discount factor, equilibrium sales in period 0 equal 6.96 and the price is 13.04. Period-1 sales equal 3.04 and the price is 16.96. The equilibrium period-1 price is a point on the dashed curve, above the intersection of the solid and the dotted curve. Discounting the future makes future revenue less valuable from the standpoint of the firm in period 0, inducing the firm to sell more in period 0. As the firm reallocates sales in this manner, the period-0 price falls, and the period-1 price rises. Equilibrium is restored when the present value of prices in the two periods are equal.

The price-taking representative firm does not shift sales from one period to another with the *intention* of causing price to change. If, following an increase of r from 0 to 0.3, the firm did not reallocate sales, then the present value of a unit of sales in period 0 remains at $p_0 = 15$ and the present value of a unit of sales in period 1 is $15\frac{1}{1.3} = 11.5 < 15$. In this case, the firm has an opportunity for intertemporal arbitrage. Prices adjust as the firm moves sales from period 1 to period 0, until, at equilibrium, there are no further opportunities for intertemporal arbitrage.

Figure 3.2 illustrates the model with constant average (= marginal) costs $c = 4$ (instead of $c = 0$ as above). The solid and dashed lines show price minus cost, instead of price, in the two periods. With zero discount rate, sales are again allocated evenly between the two periods, and the price in both periods is again $p = 15$, so price - costs = 11. In the absence of discounting, the cost increase reduces the firm's profits, but has no effect on

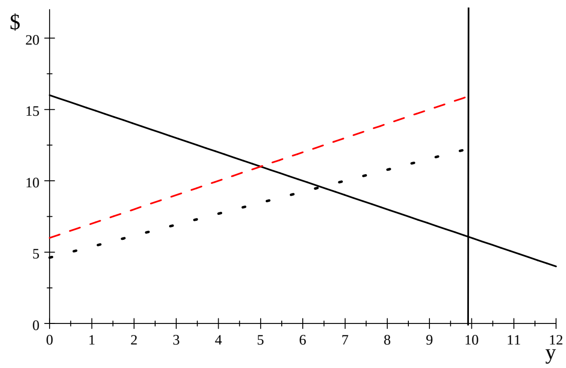

Figure 3.2: Demand - cost in period 0 (solid). Demand - cost in period 1 (dashed). Present value of period 1 price - cost (dotted).

consumers.

The dotted line in Figure 3.2 shows the present value of period-1 price minus extraction costs, for a discount factor $\rho = 0.77$. In this case, the equilibrium occurs where the present value of price minus extraction costs are equal in the two periods, at the intersection of the solid and the dotted lines. Here, period-0 sales equal 6.43 and the period-0 price is 13.57. *With discounting, higher extraction costs cause the firm to move production from period 0 to the period 1.* This reallocation causes period-0 price to rise and period-1 price to fall. With discounting, the higher extraction costs lowers consumer surplus in period 0, and increases consumer surplus in period 1.

	$\rho = 1$	$\rho = 0.77$
$c = 0$	$p_0 = 15$	$p_0 = 13.04$
$c = 4$	$p_0 = 15$	$p_0 = 13.57$

Table 3.1: Period-0 price for different discount factors and cost

Table 1 summarizes the effects of discounting and extraction costs on period-0 price. A higher extraction cost (raising costs from $c = 0$ to $c = 4$) has no effect on period-0 price in the absence of discounting, but leads to a reallocation of sales "from the present to the future" (i.e. from period 0 to period 1) under discounting. From the perspective of the firm in period 0, a one unit increase in costs increases the average and marginal extraction cost today by one unit, and increases the *present value* of cost in the next period costs by only ρ. Other things equal, higher extraction costs make it less attractive to extract the resource. However, discounting diminishes the period-1 cost-driven *disincentive* to sell, relative to the period-0 disincentive. Therefore, in the presence of discounting, firms respond to a higher cost by reducing period-0 sales and increasing period-1 sales.

Contrast the nonrenewable resource and a "standard" commodity A cost increase in a "standard" (i.e., static) supply and demand model shifts in the supply curve, resulting in a lower equilibrium supply and a higher equilibrium price. For nonrenewable resources, the effect of a cost increase depends on the discount factor. With positive discounting ($r > 0$, so $\rho < 1$), higher extraction costs cause a reallocation of supply across periods. The equilibrium price rises in one period and falls in the other. Intuition based on standard models may be misleading in nonrenewable resource markets.

3.2 Monopoly

Objectives and skills

- Write down the objective and constraints of a monopoly resource owner.
- Obtain and understand the equilibrium condition for the monopoly.
- Use graphical methods to illustrate the relation between exogenous parameters and the monopoly equilibrium.
- Compare the monopoly and the competitive outcomes (e.g. period 0 sales and price in the two cases).

The monopoly, like the competitive firm, wants to maximize the present discounted value of profits, given by expression 3.1. However, the monopoly recognizes that its sales affect the price, whereas the competitive firm takes price as given. The present discounted stream of monopoly profits equals

$$\pi^{\text{monopoly}}(y) = (p(y) - c)\, y + \rho\, (p(x - y_0) - c)(x - y)\,. \tag{3.3}$$

We can find the equilibrium condition for the monopoly by using the first order condition to the problem of maximizing $\pi^{\text{monopoly}}(y)$. A simpler approach, discussed in Chapter 2.4, finds this equilibrium condition by beginning with the equilibrium condition for the competitive industry (the last

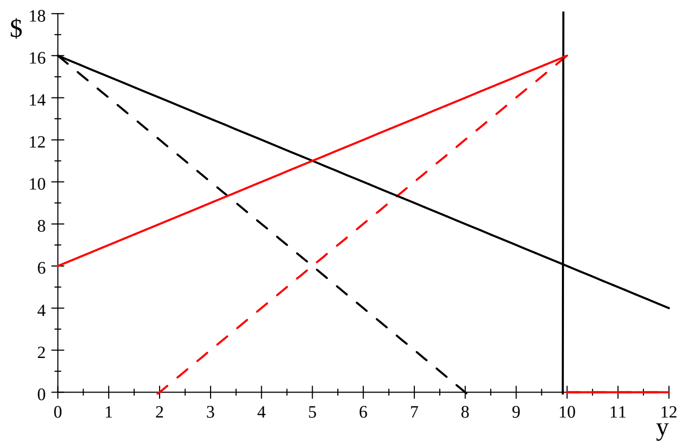

Figure 3.3: Solid curves show first and second period demand minus marginal cost as a function of first period sales. Dashed curves show marginal revenue minus marginal cost curves corresponding to these demand curves. $\rho = 1$.

part of equation 3.2), *replacing price with marginal revenue.* With sales y, marginal revenue, MR, is

$$MR(y) = p(y)\left(1 - \frac{1}{\eta(y)}\right), \text{ with } \eta(y) \equiv -\frac{dy}{dp}\frac{p}{y},$$

where η is the price elasticity of demand. The equilibrium condition for the monopoly is

$$MR(y) - c = \rho\left[MR(x - y) - c\right]. \tag{3.4}$$

Figure 3.3 shows the period-0 and period-1 demand functions, $p_0 = 20 - y$ and $p_1 = 20 - (10 - y)$, minus marginal cost, $c = 4$, (the two solid lines) as a function of period-0 sales, y. The dashed lines beneath those two demand functions show the marginal profit (marginal revenue minus marginal cost) corresponding to those two demand functions. In the absence of discounting ($\rho = 1$) optimality for the monopoly requires that marginal profit in the two periods are equal.

Absent discounting, the monopoly sells the same amount in both periods, so $y = 5$, exactly as in the competitive equilibrium with no discounting. In this market, moving from a competitive market to a monopoly has no effect on the outcome. This result is due to the fact that the stock of resource is fixed, together with the assumptions that the discount rate is 0 ($\rho = 1$) and

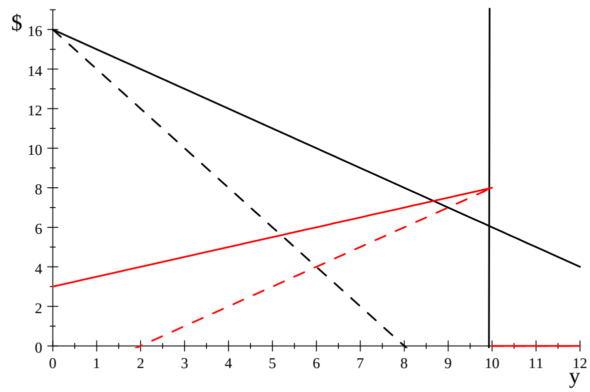

Figure 3.4: The solid curves show the present value (with $\rho = 0.5$) inverse demand functions in the two periods, minus $c = 4$, and the dashed lines show the present value marginal revenue curves, minus $c = 4$. First period sales in the competitive equilibrium $= 8.67$; first period sales under the monopoly $= 6$. "The monopoly is the conservationist's friend."

that the world lasts only two periods. Here, neither the monopoly nor the competitive firm has any incentive to save the resource beyond period 1.

The monopoly is the conservationists' friend Under our assumption that both the monopoly and competitive industry extract the same *cumulative* quantity over two periods, the two market structures lead to the same allocation of the resource if there is no discounting and if extraction costs are either 0 or constant: both sell half the aggregate quantity in each period. With discounting (or with *non-constant* average extraction costs, studied in Chapter 4), the allocations in the two equilibria are, in general, different. For the demand and cost functions used in this book (but not for all possible demand and cost combinations), the monopoly sells less in period 0, compared to the competitive industry. The monopoly therefore saves more of the resource for the future, compared to the competitive industry: "the monopoly is the conservationists' friend".

Figure 3.4 modifies Figure 3.3, including discounting with $\rho = 0.5$, illustrating that the monopoly is the conservationist's friend. Solid lines show the present value of price minus marginal cost, and dashed lines show the present value of marginal revenue minus marginal cost. The intersection of the solid lines identifies period-0 sales under competition, $y = 8.67$. The

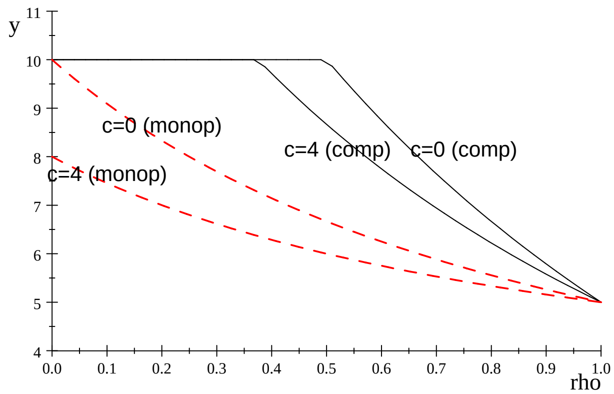

Figure 3.5: Solid curves show first period sales as a function of ρ for $c = 0$ and $c = 4$ under competition. Dashed curves show first period sales under monopoly.

intersection of the dashed lines identifies period-0 sales under the monopoly, $y = 6$. A decrease in the discount factor from $\rho = 1$ to $\rho = 0.5$ increases period-0 sales under both market structures, but the increase is greater in the competitive market. (Compare Figures 3.3 and 3.4.) A decrease in ρ shifts down and flattens both of the period-1 curves in Figure 3.4, but the marginal revenue curve is steeper than the inverse demand function, causing the point of intersection on the marginal revenue curves to move further to the left.

Figure 3.5 uses the demand function $p = 20 - Q$ and the equilibrium conditions 3.2 (for competition) and 3.4 (for monopoly) to graph period-0 sales as a function of the discount factor, ρ, for costs $c = 0$ and $c = 4$. Raising extraction costs from $c = 0$ to $c = 4$ lowers period-0 competitive sales, except for two cases: (i) when the firm discounts the future so heavily that it wants to extract everything in period 0 ($\rho < 0.4$) or (ii) when the firm does not discount the future at all ($\rho = 1$), so that it extracts the same amount in both periods. (Compare the two solid curves in Figure 3.5.) Extraction costs also reduce period-0 sales for the monopoly. The monopoly sells less in period 0 than competitive firms (for $0 < \rho < 1$). A higher valuation of the future (higher ρ, or lower r) decreases period-0 sales for both types of firm. (Appendix D provides more discussion of the comparison of the monopoly and the competitive firms.)

3.3 Comparative statics

Objectives and skills

- Reinforce the distinction between exogenous and endogenous variables, and the distinction between an explicit and implicit relation between variables.
- Answer a comparative statics question using calculus.

This section provides more practice in working through the comparative statics of a model. Following the procedure outlined in Chapter 2.2, we find an explicit expression for the endogenous variable of interest (period-0 sales) as a function of model parameters, and take derivatives to find comparative static expressions. The endogenous variables in this model are the prices and quantities in the two periods. The exogenous variables are c and ρ. A slightly richer model replaces the numerical values in the demand function with symbols, replacing $p = 20 - y$ with $p = a - by$, and replaces the initial stock, 10, with a symbol, x. For that model, the equilibrium condition in the competitive model is

$$(a - by - c) = \rho\,(a - b\,(x - y) - c)\,. \tag{3.5}$$

Because of its linearity, we can solve this equation to obtain an explicit expression for first period sales, as a function of the model parameters:

$$y = \frac{1}{b + b\rho}\,(a - c + \rho\,(c - a + bx))\,.$$

We can answer comparative statics questions by differentiating this expression with respect to model parameters:

$$\frac{dy}{dc} = \frac{\rho - 1}{b + b\rho} \leq 0 \;\text{ and }\; \frac{dy}{db} = \frac{1}{b^2}\frac{\rho - 1}{\rho + 1}\,(a - c) \leq 0. \tag{3.6}$$

The "choke price", defined as the price at which demand falls to 0, is a in this model. In order for firms to extract the resource, it must be the case that $a > c$. Therefore, the two comparative statics inequalities are "strict" ($<$ instead of $=$) for $\rho < 1$. We already showed graphically that when $\rho < 1$ an increase in extraction costs shifts extraction from the first to the second period. A larger value of b makes the inverse demand function steeper, i.e. it reduces demand at any price. The second comparative statics expression shows that this decrease in demand also reduces first period sales. (Appendix C shows how to conduct comparative statics using the implicit equation 3.5.)

3.4 Summary

Equilibrium in a nonrenewable resource market has many of the characteristics of the equilibrium in the trade example. In the two-period nonrenewable resource setting, firms can reallocate sales from one period to another. Intertemporal reallocation here corresponds to movement across space in the trade setting. An intertemporal no-arbitrage condition requires that the present value of the marginal return from selling a good is the same in both periods. The discount factor, used to convert a future receipt into its present value equivalent, plays a role analogous to transportation costs in the trade model. Transportation costs in the trade model cause prices to differ between the two locations. Positive discounting ($r > 0$, $\rho < 1$) causes prices to differ across periods in the resource setting.

We obtained the equilibrium condition for a monopoly by taking the equilibrium condition for a competitive firm, and replacing price with marginal revenue, where marginal revenue $= p\,(1 - 1/\eta)$, and η is the price elasticity of demand. With constant marginal extraction costs and no discounting ($r = 0$, so $\rho = 1$), both types of firms sell the same amount, half of the available stock, in period 0. Under discounting, ($\rho < 1$), the monopoly sells less in the first period than the competitive firm. Here, the monopoly is the conservationist's friend.

Graphical methods show that for $\rho < 1$, higher extraction costs lower period-0 sales under both competition or monopoly. Higher costs decrease the sales incentive in both periods; but because of discounting, the incentive for period-1 sales (= extraction) falls by less than does the incentive for period-0 sales. We also used calculus to answer comparative statics questions.

3.5 Terms, study questions, and exercises

Terms and concepts Extraction costs, intertemporal arbitrage, trajectory, "monopoly is the conservationist's friend", choke price.

Study questions For these questions, use the linear inverse demand function, $p = 10 - y$.

1. In the two-period setting, with discount factor $\rho < 1$, use a figure to describe the effect of an increase in extraction costs, from $C = 0$ to

$C = 2$, on the equilibrium price and sales trajectory. Provide the economic explanation for this change.

2. In a two period setting with linear demand and constant average extraction costs C use two figures to illustrate the equilibrium sales trajectory under competition and under monopoly for the two cases where (a) the discount factor is $\rho = 1$ and (b) the discount factor is less than 1. Explain the effect of discounting.

3. Answer questions 1 and 2 algebraically, using the equilibrium conditions under competition and monopoly.

Exercises

1. Find the first order condition to the problem of maximizing $\pi^{\text{monopoly}}(y)$ and show that this first order condition is identical to equation 3.4.

2. Assume that (for whatever reason) the resource can be extracted during only two periods. (After the second period, any remaining stock is worthless.) (a) Using the demand function $p = 20 - y$, constant extraction costs $c = 5$, and a discount factor ρ, find the critical level of the initial stock, x^c (a number) such that a competitive equilibrium exhausts the resource if and only if the initial stock, x, satisfies $x \leq x^c$. (b) If $x \geq x^c$, what is true of price in the two periods? (c) Does the critical value depend on ρ? Does it depend on c? Explain.

3. In the two-period model, suppose that $p = 20 - q$, $c = 5$, initial stock $x = 5$. Find the critical value of ρ, call it ρ^{crit}, such that period 1 extraction is 0 for $\rho \leq \rho^{crit}$. Provide an economic explanation for this possibility.

4. Suppose that $c = 0$, the discount factor is ρ, and demand is constant elasticity, $y = p^{-\eta}$. (a) Write the equilibrium conditions for the competitive firm and the monopoly in this case. (b) In order for the monopoly equilibrium condition to be sensible, what restriction must be imposed on η? Provide the economic explanation for this restriction. (c) Compare the level of first period sales under competition and under monopoly.

5. Consider the two-period model with inverse demand $p = a - bq$, constant average extraction cost c, initial stock x, and discount factor ρ. Period 0 extraction is y, the endogenous variable. Suppose that in the competitive equilibrium extraction is positive in both periods, and the resource constraint is binding. Find $\frac{dy}{d\rho}$ and give the economic explanation for the sign of this derivative (one or two sentences). (Hint: Use Section 3.3 to find the expression for y as a function of ρ and other model parameters. No need to re-derive this function. Take the derivative of this function with respect to ρ to find $\frac{dy}{d\rho}$. You will discover that the sign of this derivative depends on the sign of $(2c - 2a + bx)$. The trick is to determine the sign of this expression. Proceed as follows. Find the level of sales in both periods if the resource constraint is *not* binding. The sum of these two sales levels gives a critical initial stock level: the initial stock must be at least at this critical level, if the resource constraint is not binding. Because you are told that the resource constraint *is* binding, you know that the initial stock must be below this critical level. This information enables you to determine the sign of $(2c - 2a + bx)$ and thereby determine the sign of $\frac{dy}{d\rho}$.

6. The text assumes that extraction is positive in both periods. Using the demand and cost assumptions in the example in Chapter 3.1, find the critical discount factor, denoted ρ^*, such that second period sales in the competitive equilibrium are 0 if $\rho < \rho^*$. Provide the economic intuition.

Chapter 4

Additional tools

Objectives

- Work with a stock-dependent cost function; use the "perturbation method" to obtain equilibrium conditions; and express these conditions using "rent".

Information and skills

- Understand the rationale for using a stock-dependent extraction cost function, and be able to work with a particular cost function.
- Write down a firm's objective function and constraints.
- Derive and interpret the optimality condition to this problem, for the two cases where the resource constraint is binding or is not binding.
- Understand the logic of the perturbation method, and apply it in the two-period setting.
- Understand the meaning of "rent" in the resource setting, and use it to express the optimality (equilibrium) condition.
- Understand the relation between rent in period 0 and in period 1.

We build on the previous chapter, introducing: (i) a more general cost function, (ii) the "perturbation method", and (iii) the concept of rent. The constant-average-cost model in Chapter 3 provides intuition, but obscures important features of many resource settings. By considering a more general cost function at the outset, we can present the subsequent material concisely, without repeating steps for each important special case.

The "perturbation method" provides a quick way to obtain the equilibrium condition in resource models. The idea behind this method, if not the term, will be familiar to many readers. Imagine the firm beginning with a "candidate" for an optimal plan, e.g. selling 53% in period 0 and 47% in period 1. The firm can test whether this candidate is optimal by "perturbing" it, moving a small (infinitesimal) amount of sales from one period to the other. If this perturbation increases the firm's present discounted value of profits, the original candidate was not optimal. It the perturbation decreases the firm's profits, then using the "opposite" perturbation, e.g. moving sales from period 0 to period 1, instead of from period 1 to period 0, would increase profits. Thus, if the perturbation either increases or decreases profits, the candidate is not optimal. In order for the candidate to be optimal, an infinitesimal perturbation must have *"zero first order effect"* on the payoff. This statement means that *the derivative of the payoff, with respect to the perturbation, evaluated at a zero perturbation, is zero.*

The concept of rent is important in resource economics. "Rent" is a common word, but it has a particular meaning in economics, and a still more particular meaning in resource economics. It provides a convenient way to express the equilibrium conditions. This chapter considers only the competitive equilibrium. Rather than duplicate the analysis for the monopoly, we merely note that replacing "price" with "marginal revenue" in the competitive condition, yields the equilibrium condition for monopoly. Numerical examples, collected in Chapter 4.5, illustrate the methods.

4.1 A more general cost function

Objectives and skills

- Understand the reasons for allowing average extraction costs to depend on the stock and the extraction level.
- Understand the relation between parameter values and the characteristics of cost for an example.

The distinction between stock and flow variables is central to resource economics. A stock variable is measured in units of quantity, e.g. billions of barrels of oil, or tons of coal, or number of fish, or gigatons of carbon. The units of measurement do not depend on units of time. The number of tons of coal might, of course, change over time, but the statement that we have x tons of coal today does not depend on whether we measure time in months or years. In contrast, the units of measurement of flow variables do depend on units of time. For example, the statement "This well produces 1000 barrels of oil" is meaningless unless we know whether it produces this number of barrels per hour, day, or week. The variable x_t denotes the stock of a resource, with the subscript identifying time, or the period number. The variable y_t is a flow variable, denoting extraction during a period. If a period lasts for one year, and quantity units are tons, then x_t is in units of tons and and y_t is in units of *tons per year*.

The constant average cost function used in Chapter 3 assumes that marginal extraction costs do not depend on either the size of the remaining resource stock or on the rate of extraction. We relax both of these assumption. Marginal (and average) extraction costs typically increase as the size of the remaining stock falls. This relation likely holds at both the level of the individual mine or well, and at the economy-wide level. At the individual level, shallow and relatively inexpensive wells are adequate to extract oil or water when the stock of oil in a field or water in an aquifer is high. As the stocks diminish, it becomes necessary to dig deeper and more expensive wells to continue extraction. At the economy-wide level, different deposits have different extraction costs. Because it is (generally) efficient to extract from the cheaper deposits first, extraction costs increase as the size of the remaining economy-wide stock falls. People began mining coal from seams that lay close to the ground; early oil deposits could be scooped up with little effort. As society exhausted these cheap deposits, it became economical to remove mountaintops to obtain coal and to exploit deep-water deposits to extract oil. Extraction costs rose as remaining economy-wide resource stocks fell.[1]

If the rate of extraction does not affect average and marginal cost implies, then total extraction costs double in a period if we double the amount extracted. In many circumstances, average and marginal costs increase with the rate of extraction. For example, it might be necessary to pay workers overtime or to hire less qualified workers in order to increase extraction in

[1] Many resource firms are vertically integrated, both extracting and processing natural resources. Some of the empirical literature distinguishes between extraction and processing costs.

a period. In this situation average and marginal extraction costs increase with the extraction rate.

It is important not to confuse stock-dependent extraction costs with increasing average and marginal costs. The former causes average or marginal costs to rise over time, as the stock falls; the latter causes higher extraction within a period to increase these costs. Average and marginal extraction costs might increase for either or both of these reasons, but the two types of cost-related considerations are distinct. To accommodate both of these features, we need a (total) cost function of the form $c(x,y)$, with the following characteristics

$$\frac{\partial c(x,y)}{\partial x} \leq 0, \quad \frac{\partial \left[\frac{c(x,y)}{y}\right]}{\partial y} \geq 0, \quad \frac{\partial^2 c(x,y)}{\partial y^2} \geq 0.$$

The first inequality states that a higher stock either lowers costs or (in the case of equality) leaves them unchanged. The second states that higher extraction either increases *average* costs or leaves them unchanged. The third states that higher extraction either increases *marginal* costs or leaves them unchanged.

A parametric example makes this cost function concrete:

$$\text{Parametric example: } \; c(x,y) = C(\sigma + x)^{-\alpha} y^{1+\beta}, \tag{4.1}$$

where C, α, σ, and β are non-negative parameters. Table 1 shows the relation between parameter values and marginal costs.

parameter values	cost function	marginal cost	marginal extraction costs are:
$C = 0$	0	0	zero
$C > 0$, $\alpha > 0$ and $\beta > 0$	$C(x+\sigma)^{-\alpha}y^{1+\beta}$	$C(1+\beta)(x+\sigma)^{-\alpha}y^{\beta}$	increasing in extraction, decreasing in stock
$C > 0$, $\alpha = 0$ and $\beta = 0$	Cy	C	independent of both extraction and stock
$C > 0$, $\alpha = 0$, and $\beta > 0$	$Cy^{1+\beta}$	$C(1+\beta)y^{\beta}$	increasing in extraction, independent of stock
$C > 0$, $\alpha > 0$, and $\beta = 0$	$C(x+\sigma)^{-\alpha}y$	$C(x+\sigma)^{-\alpha}$	independent of extraction, decreasing in the stock

Table 4.1: Relation between parameter values and marginal extraction costs.

A caveat This model uses a single stock variable, x, and ignores the discovery and development of new stocks. Here, the resource stock falls over time, with extraction. In fact, new discoveries frequently occur, raising the size of "proven reserves" (known stocks) (Chapter 6.3). Stocks with cheaper extraction costs tend to be used first (Chapter 5.5), and newly discovered stocks are often of lower quality, i.e., have higher extraction costs. If we treated new discoveries as an increase in x, then our model of extraction costs would suggest (incorrectly) that these discoveries tend to reduce extraction costs. However, the discovery of new stocks, e.g. in the Aortic region, obviously do not decrease the cost of extracting Saudi oil.

4.2 The perturbation method

Objectives and skills

- Write down the firm's objectives and constraints, based on a statement of the problem.
- Write down and interpret the first order condition (= optimality condition) for this problem, both in the case where the resource constraint is binding and where it is not binding (= "slack").
- Review the "standard" method of obtaining the optimality condition, and introduce the perturbation method.

This section uses two approaches to derive the necessary condition for optimality (the "equilibrium condition") in the two-period competitive market. The standard approach begins by (i) eliminating the constraint by substitution, (ii) then taking the derivative of the present discounted value of profits with respect to period-0 sales, and (iii) finally, replacing the price in each period (which the firm takes as exogenous) with the inverse demand function. The second approach uses the perturbation method. The perturbation method is useful for models with many periods, so we introduce it in a setting where it is easier to understand.

The initial stock of the resource, at the beginning of period 0, is x_0. A candidate consists of feasible extraction levels in the two periods, y_0 and y_1. These must satisfy the resource constraint and the non-negativity constraints:

$$0 \leq y_0 \leq x_0,\ 0 \leq y_1 \leq x_1 = x_0 - y_0 \text{ and } x_1 - y_1 \geq 0.$$

Extraction cannot be negative, and cannot exceed available stock; the available stock in period 1 equals the initial stock minus the amount that was extracted in period 0. The final inequality states that the ending stock, after period-1 extraction, must be non-negative. If extraction is positive in both periods, and all of the resource is used, the constraints imply

$$y_1 = x_1 = x_0 - y_0. \tag{4.2}$$

There are some circumstances where it is not optimal to use all of the resource; in that case, $y_1 < x_1$ instead of $y_1 = x_1$.

To understand these two cases, it helps to consider the firm's problem at period 1, after it has already made the period-0 extraction decision. In period 1, the firm has the remaining stock $x_1 = x_0 - y_0$. The firm can extract all the stock or leave some in the ground. Equation 4.3 summarizes the second period extraction rule:

$$\left(p_1 - \frac{\partial c\left(x_1, y_1\right)}{\partial y_1}\right)_{|y_1 = x_1} \begin{cases} \geq 0 \Leftrightarrow y_1 = x_1 \\ < 0 \Leftrightarrow y_1 < x_1 \end{cases} \tag{4.3}$$

The first inequality states that if extracting everything (evaluation the derivative at $y_1 = x_1$) leads to a price greater than or equal to marginal cost, then the firm does want to extract everything. The second inequality states that if extracting everything leads to price less than marginal cost, then the firm leaves some stock in the ground (so $y_1 < x_1$).

4.2.1 It is optimal to use all of the resource

Here we assume that in equilibrium $y_1 = x_1$; using the first line of equation 4.3, this assumption implies that marginal profit at period 1 is greater than or equal to zero. The present discounted value of total profit for the price-taking firm is

$$p_0 y_0 - c\left(x_0, y_0\right) + \rho\left[p_1 y_1 - c\left(x_1, y_1\right)\right]. \tag{4.4}$$

The "standard" method of obtaining equilibrium condition. We can substitute the constraints 4.2 into the objective, to write the present discounted value of profits as

$$\pi\left(y_0\right) = p_0 y_0 - c\left(x_0, y_0\right) + \rho\left[p_1\left(x_0 - y_0\right) - c\left(x_0 - y_0, x_0 - y_0\right)\right].$$

The first order condition to the problem of maximizing $\pi(y_0)$ is

$$\frac{d\pi(y_0)}{dy_0} \stackrel{\text{set}}{=} 0.$$

This first order condition implies (Box 4.1):

$$p_0 - \frac{\partial c(x_0, y_0)}{\partial y_0} = \rho \left[\underline{\left(p_1 - \frac{\partial c(x_1, y_1)}{\partial y_1} \right)} \underbrace{- \frac{\partial c(x_1, y_1)}{\partial x_1}} \right] \tag{4.5}$$

The optimal decision balances the gain from additional extraction in period 0 (the left side of equation 4.5) with the loss from lower extraction and higher costs in period 1 (the right side of the equation). The left side is the familiar "price minus marginal cost", the increase in period-0 profits from extracting one more unit in that period. The right side is the present value of two terms, the underlined and the "under-bracketed" terms. The underlined term equals the reduction in period-1 profit, the loss arising from having one less unit to sell. The under-bracketed term is the cost increase due to a reduction in the stock at the beginning of period 1, resulting from the higher period-0 extraction. If costs are independent of the stock, then $\frac{\partial c(x_1,y_1)}{\partial x_1} = 0$; in this special case, the under-bracketed term vanishes and equation 4.5 then states that the present value of marginal profit is equal in periods 0 and 1.

Box 4.1 Derivation of equation 4.5 The first order condition for the competitive firm's maximization problem is

$$\frac{d\pi(y_0)}{dy_0} = \left[p_0 - \frac{\partial c(x_0,y_0)}{\partial y_0} \right] +$$

$$\rho \left[p_1 \frac{dy_1}{dy_0} - \frac{\partial c(x_1,y_1)}{\partial x_1} \frac{dx_1}{dy_0} - \frac{\partial c(x_1,y_1)}{\partial y_1} \frac{dy_1}{dy_0} \right] \stackrel{set}{=} 0.$$

The second line uses the chain rule. For example, period-1 costs depend on the period-1 stock; $\frac{\partial c(x_1,y_1)}{\partial x_1}$ picks up this dependence. From the first constraint in equation 4.2, the period-1 stock depends on period-0 extraction, via $\frac{dx_1}{dy_0} = -1$. Similarly, $\frac{dy_1}{dy_0} = -1$.
We can use these two equalities to write the first order condition as

$$\frac{d\pi(y_0)}{dy_0} = \left[p_0 - \frac{\partial c(x_0, y_0)}{\partial y_0} \right] + \rho \left[-p_1 + \frac{\partial c(x_1, y_1)}{\partial x_1} + \frac{\partial c(x_1, y_1)}{\partial y_1} \right] \stackrel{set}{=} 0.$$

Rearranging this condition gives equation 4.5.

The alternative: perturbation method The approach used above to derive equation 4.5 is cumbersome in the many-period problem. With that problem in mind, we consider the method of perturbation: a different route to the same goal. We begin with a candidate, y_0 and y_1, and the associated period-1 stock, $x_1 = x_0 - y_0$. The assumption that it is optimal to consume all of the resource, means that any candidate worth considering sets $y_1 = x_1$. Expression 4.4 shows the payoff associated with this candidate.

We can "perturb" this candidate by changing period-0 extraction by a small (positive or negative) amount, ε. Because, (by assumption) it is optimal to consume all of the resource, a change in period-0 extraction of ε requires an offsetting change in period-1 extraction of $-\varepsilon$. The "gain" from a perturbation, $g(\varepsilon; y_0, x_1, y_1)$, is

$$\begin{gathered} g(\varepsilon; \cdot) = p_0 \times (y_0 + \varepsilon) - c(x_0, y_0 + \varepsilon) + \\ \rho \left[p_1 \times (y_1 - \varepsilon) - c(x_1 - \varepsilon, y_1 - \varepsilon) \right]. \end{gathered} \tag{4.6}$$

If the candidate is optimal, then a perturbation causes zero first order change to the payoff: at an optimum

$$\frac{dg(\varepsilon; y_0, x_1, y_1)}{d\varepsilon}_{|\varepsilon=0} = 0. \tag{4.7}$$

Evaluating this derivative (Box 4.2) produces the same first order condition obtained above, equation 4.5.

4.2.2 It is optimal to leave some of the resource behind

Important resources, e.g. coal, are unlikely to be physically exhausted; at some point, remaining deposits become too expensive to extract and are left in the ground. Here we consider the situation where it is optimal to *not* exhaust the resource: $y_1 < x_1$, i.e. the resource constraint is "slack".

The firm stops extracting before marginal profits become negative. With period-1 profits $p_1 y_1 - c(x_1, y_1)$, marginal profits equal price minus marginal cost, the underlined term on the right side of equation equation 4.5. The firm does not want to exhaust the stock if doing so creates negative marginal profits, as the second line of equation 4.3 states. If the marginal profit of extracting the last unit is negative, then the firm extracts up to the point where period-1 marginal profit is 0:

$$p_1 - \frac{\partial c(x_1, y_1)}{\partial y_1} = 0. \tag{4.8}$$

Using this equality in condition 4.5 (setting the underlined term equal to 0) yields the equilibrium condition.

$$p_0 - \frac{\partial c(x_0, y_0)}{\partial y_0} = -\rho \left[\frac{\partial c(x_1, y_1)}{\partial x_1} \right]. \tag{4.9}$$

Box 4.2 Evaluating the derivative in equation 4.7. Because ε appears in two places in the period-1 cost function, $c(x_1 - \varepsilon, y_1 - \varepsilon)$, we use the *total* derivative to evaluate the effect of ε on this function. Using the chain rule and

$$\frac{d(x_1 - \varepsilon)}{d\varepsilon} = \frac{d(y_1 - \varepsilon)}{d\varepsilon} = -1,$$

the total derivative of period-1 costs, with respect to ε, *evaluated at* $\varepsilon = 0$ is:

$$\frac{dc(x_1 - \varepsilon, y_1 - \varepsilon)}{d\varepsilon}_{|\varepsilon=0} = -\left(\frac{\partial c(x_1, y_1)}{\partial x_1} + \frac{\partial c(x_1, y_1)}{\partial y_1} \right).$$

Using this equation, we have

$$\frac{dg(\varepsilon; y_0, x_1, y_1)}{d\varepsilon}_{|\varepsilon=0} = p_0 - \frac{\partial c(x_0, y_0)}{\partial y_0} - \rho \left[p_1 - \left(\frac{\partial c(x_1, y_1)}{\partial x_1} + \frac{\partial c(x_1, y_1)}{\partial y_1} \right) \right].$$

Set this derivative to 0 and rearrange to obtain condition 4.5.

4.3 Solving for the equilibrium

Objectives and skills

- Know how to use the optimality condition and the constraints to solve for the equilibrium prices and sales levels.

Firms take prices as exogenous, but they are determined by equilibrium behavior. (Prices are "endogenous to the model" – not to the firm.) Given specific demand and cost functions, we have enough information to actually solve for the equilibrium. This model contains three endogenous variables,

y_0, y_1 and x_1. We need three equations to find these three variables. The trick is to identify these three equations and then know how to use them. We describe the process here, and illustrate it in Chapter 4.5 using examples.

We have to consider three cases: (i) the resource might be exhausted in period 0, leaving nothing to extract in period 1 ($y_0 = x_0$, so $y_1 = 0$); (ii) the resource might be exhausted in period 1, with positive extraction in both periods ($0 < y_0 < x_0$ and $y_1 = x_0 - y_0$); (iii) the resource might not be exhausted ($y_0 + y_1 < x_0$). We proceed as follows. First, solve the model under the *assumption* that we are in case (ii). Second, determine whether the assumption is correct.

- Step 1. In all of these cases, the constraint $x_1 = x_0 - y_0$ provides one of the three equations; we need two more equations. The assumption that we are in Case (ii) implies $y_1 = x_0 - y_0$; the necessary condition 4.5 is the third equation. We solve these to obtain y_0, which gives x_1 and y_1.

- Step 2 If our solution from Step 1 gives $y_0 > x_0$, then $x_1 < 0$, violating the non-negativity constraint. In this situation, we conclude that the non-negativity constraint is binding, so $y_0 = x_0$, implying that $y_1 = 0$. Here, all of the resource is used during period-0: we are in Case (i).

- Step 3 If our proposed solution from Step 1 satisfies $y_0 < x_0$, then we check whether rent in period 1 is non-negative. If $R_1 \geq 0$, our proposed solution is correct. If, however, the proposed solution from Step 1 implies that $R_1 < 0$, that solution is incorrect. We have now ruled out both Cases (i) and (ii), so we conclude that Case (iii) is correct. Our three equations consist of the constraint, $x_1 = x_0 - y_0$, the necessary condition 4.9, and equation 4.8. We solve these equations to find y_0 and y_1.

4.4 Rent

Objectives and skills

- Know the meaning of rent; write the optimality condition using rent.

- Understand the relation between rent in the two periods.

- Understand why period-0 rent depends on whether the resource is exhausted, and on whether extraction costs depend on the stock.

"Rent" is the payment to a factor of production that exceeds the amount necessary to make that factor available. The classic example is rent to *unimproved* land used in production. Because the land is in limited supply, it receives a payment. The payment is not needed in order to create the land – it already exists, regardless of the payment. However, the limited supply means that other potential users are willing to bid for the land; the highest value use determines the rent in a competitive market. Natural resources, like land, are limited, so they command rent.

Very little productive land is "unimproved". Usually, a previous investment increased the land's productivity by, for example, removing trees and rocks. Once these improvements have been made, they are sunk, but they still receive a payment because of their limited supply. The improvements continue to exist regardless of whether the payment is actually made. Because the *actual* payments are not necessary for the continued existence of the improvements, they resemble rent. However, the improvements were made with the *anticipation* of the payments, so the payments are not precisely rent. For this reason, payments resulting from a sunk investment are known as "quasi-rent". Most natural resource stocks become available only after significant investments in exploration and development. Thus, the payments in excess of extraction costs, arising from the sale of resources, are the sum of rents and quasi-rents. Until Chapter 11 we ignore this distinction, and refer the resource rent merely as "rent".

In competitive markets (resource) rent is defined as the difference between price and marginal extraction costs. Denoting rent in period t as R_t, we have

$$R_0 = p_0 - \frac{\partial c\left(x_0, y_0\right)}{\partial y_0} \text{ and } R_1 = p_1 - \frac{\partial c\left(x_1, y_1\right)}{\partial y_1}.$$

We can use this definition to write the optimality conditions in the two cases where the resource is exhausted or is not exhausted, using a single equation. If the resource is exhausted, then $R_1 \geq 0$; except for knife-edge cases, the inequality is strict. If the resource is not exhausted, then $R_1 = 0$. We can write the equilibrium conditions 4.5 and 4.9 as[2]

$$R_0 = \rho \left(R_1 - \frac{\partial c (x_1, y_1)}{\partial x_1} \right). \tag{4.10}$$

Equation 4.10 states that period-0 rent equals the present value of the sum of two terms: period-1 rent, plus the cost reduction due to having higher period-1 stock. Either of the terms on the right side could be zero or positive (but never negative). These two terms capture the two reasons that period-0 rent is (typically) positive:

1. Scarcity: we will run out of the resource. Extracting one more unit today means that we have one less unit to extract in the future. That extra unit of potential future extraction is valuable if and only if $R_1 > 0$. If, instead, $R_1 = 0$, then we will *not* run out of the resource (the resource is not scarce), thus eliminating one of the reasons that period-0 rent is positive.

2. Stock-dependent extraction costs: extraction of an extra unit today makes future extraction more expensive. If, however $-\frac{\partial c(x_1,y_1)}{\partial x_1} = 0$ (extraction cost does not depend on stock) this reason for positive period-0 rent also vanishes.

4.5 Examples

Four examples illustrate the methods developed above. Examples 1 and 2 illustrate the perturbation method in the two cases where the firm either does or does not exhaust the resource. Examples 3 and 4 show how to solve the equilibrium when marginal extraction costs are either constant or decreasing in the stock.

Example 1 This example illustrates the perturbation method for inverse demand $p = a - by$, initial stock x_0, discount factor ρ and extraction cost function $c(x, y) = \frac{y}{10+x}$, in the situation where the firm exhausts the

[2] Under monopoly, we define rent as *marginal revenue* (instead of price) minus marginal cost. With this modification, equation 4.10 gives the monopoly equilibrium condition.

resource ($y_1 = x_1$). The competitive firm's objective and constraints are

$$\begin{array}{c} \max_{y_0,y_1,x_1} p_0 y_0 - \frac{y_0}{10+x_0} + \rho\left[p_1 y_1 - \frac{y_1}{10+x_1}\right] \\ \text{subject to: } x_1 = x_0 - y_0 \text{ and } y_0 \geq 0, x_1 \geq 0, x_1 - y_1 \geq 0. \end{array} \tag{4.11}$$

We write the objective using the prices, p_0 and p_1, *not the inverse demand function*, reflecting the fact that the competitive firm takes prices as given. A "candidate" consists of values of y_0, x_1, and y_1 that satisfy the constraints (i.e. are feasible). A perturbation changes y_0 to $y_0 + \varepsilon$ and changes x_1 to $x_1 - \varepsilon$: if the firm extracts ε more units in period 0, the stock remaining at period 1 is reduced (relative to the candidate) by ε. This example assumes that the candidate exhausts the resource, so the perturbation changes y_1 to $y_1 - \varepsilon$. For example, if $\varepsilon > 0$, then the perturbation reduces x_1 by a small amount, making it necessary to reduce y_1 by an equal amount in order to satisfy the non-negativity constraint on end-of-period stock. The gain function is

$$g(\varepsilon;\cdot) = p_0 \times (y_0 + \varepsilon) - \frac{y_0 + \varepsilon}{10 + x_0} + \rho\left[p_1 \times (y_1 - \varepsilon) - \frac{y_1 - \varepsilon}{10 + x_1 - \varepsilon}\right] \tag{4.12}$$

and the necessary condition is

$$\frac{dg(\varepsilon; y_0, x_1, y_1)}{d\varepsilon}_{|\varepsilon=0} = \left[p_0 - \frac{1}{10 + x_0}\right] + \rho\left[-\left(p_1 - \frac{1}{10 + x_1}\right) - \frac{y_1}{(10 + x_1)^2}\right] \overset{\text{set}}{=} 0.$$

We can rearrange the last equation to write the necessary condition as

$$p_0 - \frac{1}{10 + x_0} = \rho\left[\left(p_1 - \frac{1}{10 + x_1}\right) + \frac{y_1}{(10 + x_1)^2}\right].$$

In order to find the equilibrium values, y_0, y_1, x_1, we replace price by the inverse demand function and use the constraints to obtain an equation for the endogenous y_0 as a function of the exogenous initial stock, x_0.

$$\begin{array}{c} a - by_0 - \frac{1}{10+x_0} = \\ \rho\left[a - b(x_0 - y_0) - \frac{1}{10+x_0-y_0} + \frac{x_0-y_0}{(10+x_0-y_0)^2}\right]. \end{array} \tag{4.13}$$

Example 2 This example illustrates the perturbation method when the firm does not exhaust the resource: $x_1 - y_1 > 0$. Here, a small increase in y_0 (leading to a decrease in x_1) does not require a reduction in y_1. We use the demand and cost functions from Example 1. Equation 4.11 shows the firm's objectives and constraints. The gain function is

$$g(\varepsilon;\cdot) = p_0 \times (y_0 + \varepsilon) - \frac{y_0 + \varepsilon}{10 + x_0} + \rho \left[p_1 \times y_1 - \frac{y_1}{10 + x_1 - \varepsilon} \right]. \tag{4.14}$$

The gain functions in equations 4.12 and 4.14 differ only in the term in square brackets. The former involves $y_1 - \varepsilon$ (reflecting the fact that the changed extraction in period 0 requires an offsetting change in period 1) and the latter involves y_1 (reflecting the fact that a change in period-0 extraction does not require an offsetting change in period 1). The necessary condition for optimality is

$$\frac{dg(\varepsilon; y_0, x_1, y_1)}{d\varepsilon}_{|\varepsilon=0} = \left[p_0 - \frac{1}{10 + x_0} \right] - \rho \left[\frac{y_1}{(10 + x_1)^2} \right] \overset{\text{set}}{=} 0.$$

In order to find the equilibrium values, we replace the price by the inverse demand function and use the constraints to obtain an equation for y_0:

$$a - by_0 - \frac{1}{10 + x_0} = \rho \left[\frac{y_1}{(10 + x_0 - y_0)^2} \right]. \tag{4.15}$$

Example 3 This example uses the definition of rent and the explanation in Chapter 4.3 to show how to obtain the equilibrium in the case of linear demand and constant marginal extraction costs, Cy, using $C = 4$, $\rho = 0.77$ and $x_0 = 10$. When extraction costs do not depend on the stock, equation 4.10 simplifies to $R_0 = \rho R_1$: the present value of rent is the same in both periods. In the interest of brevity, we ignore the possibility that all of the resource is consumed in period 0, leaving two remaining possibilities: the resource is exhausted over two periods, or the resource is not exhausted (Cases ii and iii from Chapter 4.3). We consider a high demand ($p = 20 - y$) and a low demand ($p = 7 - y$) scenario, in order to illustrate these two possibilities, and also to show how to compute the equilibrium.

We begin by using the equilibrium condition 4.10 under the assumption that the resource is exhausted. Our three equations are: the optimality condition $R_0 = \rho R_1$; the constraint $x_1 = x_0 - y_0$; and the assumption that we exhaust the resource, $y_1 = x_1$. We use the last two equations to write $y_1 = x_0 - y_0$. Substituting this equation into $R_0 = \rho R_1$ gives, for the high-demand scenario

$$\begin{aligned} 20 - y_0 - 4 &= 0.77\,(20 - (10 - y_0) - 4) \Rightarrow \\ y_0 &= 6.4 \Rightarrow R_0 = 20 - 6.4 - 4 = 9.6 > 0. \end{aligned}$$

Because the present value of rent is the same in both periods, we know that $R_1 > 0$. Thus, $y_0 = 6.4$ and $y_1 = 10 - 6.4 = 3.6$ is the equilibrium for this problem.

In the low demand scenario, $R_0 = \rho R_1$ implies

$$\begin{aligned} 7 - y_0 - 4 &= 0.77\,(7 - (10 - y_0) - 4) \Rightarrow \\ y_0 &= 4.74 \Rightarrow R_0 = 7 - 4.74 - 4 = -1.74 < 0. \end{aligned}$$

Here, the assumption that the resource is exhausted implies that rent is negative. Firms do not loose money, so the assumption must be false. In this case, we know that the firm does not exhaust the resource, so its period-1 rent is zero. With stock-independent extraction costs, the present value of rent is the same in both periods. Therefore, we know that period-0 rent is also 0. Thus, equilibrium requires $7 - y - 4 = 0$, or $y = 3$ in both periods.

Figure 4.1 illustrates these two possibilities; review Figure 3.1 if Figure 4.1 is unclear. The solid lines in this figure show the present discounted value of price minus marginal cost with high demand and the dashed lines show these relations with low demand. Under the assumption that the resource is exhausted, the equilibrium occurs at the intersection of the (solid or dashed) curves. This intersection lies above the y axis, i.e. it corresponds to positive rent = price – marginal cost in the high demand scenario. There, the intersection gives the equilibrium. The intersection lies below the y axis, i.e. it corresponds to negative rent in the low demand scenario. There, the intersection does not give the equilibrium (because rent is never negative). Consequently, the equilibrium occurs where the dashed lines intersect the y axis: extraction is 3 in both periods.

Example 4 Here we consider the more complex situation, where extraction costs depend on the resource stock. We are not able to obtain the equilibrium in closed form. However, deriving the equilibrium conditions provides practice in working with this model, and makes it possible to obtain a numerical solution.

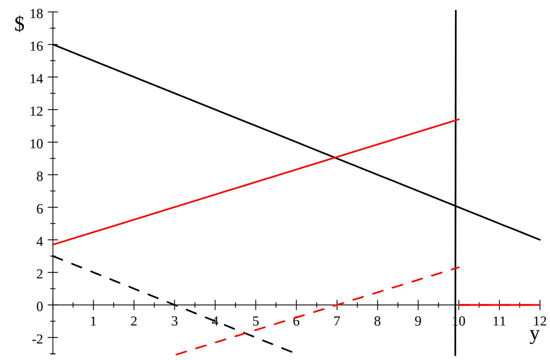

Figure 4.1: Solid lines shows price minus marginal cost with high demand $p = 20 - y$, where the resource is exhausted. Dashed lines show price minus marginal cost with low demand $p = 7 - y$, where the resource is not exhausted.

We use the cost function in Example 1, and set inverse demand to $p = 10 - y$, with $\rho = 0.77$. We leave x_0 as a free parameter, in order to show how the solution depends on the initial stock. Following Step 1 from Chapter 4.3, we first solve for the equilibrium under the assumption that extraction is positive in both periods and it is optimal to exhaust the stock. Our three equations are: the optimality condition 4.13; the constraint, $x_1 = x_0 - y_0$; and the assumption that we exhaust the resource, $y_1 = x_1$. Using the second two equations, we can write the optimality condition as

$$10 - y_0 - \frac{1}{10 + x_0} = 0.77 \left[10 - (x_0 - y_0) - \frac{1}{10 + x_0 - y_0} + \frac{x_0 - y_0}{(10 + x_0 - y_0)^2} \right].$$

The solid graph in Figure 4.2 shows the solution to this equation, y_0 as a function of $x_0 \in [0, 12]$.

We now proceed to Step 2. The dashed graph shows the 45^o line. Comparison of the solid and the dashed graph shows that our assumption implies $y_0 > x_0$, i.e. the non-negativity constraint is violated, whenever x_0 is less than 2.3. Thus, we know that for $x_0 \leq 2.3$ it is optimal to extract all of the resource in period 0 (Case (i)). For $x_0 > 2.3$ we are either in Case (ii) (as our assumption claims) or Case (iii).

We now proceed to Step 3. If the initial stock is extremely large, it is not optimal to exhaust the stock (in our two-period setting). We can solve

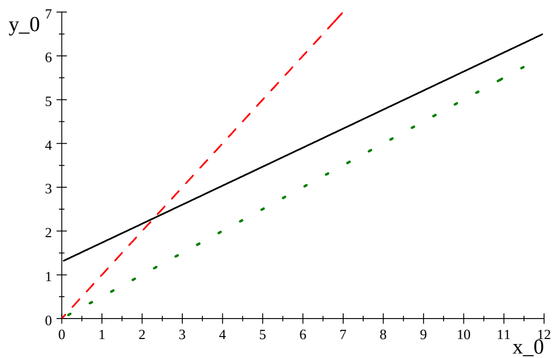

Figure 4.2: The solid graph shows y_0 as a function of the initial stock, x_0, under the assumption that extraction is positive in both periods, and all of the resource is extracted.. The dashed line shows the graph of $y_0 = x_0$ and the dotted line shows the graph of $y_0 = 0.5x_0$.

$R_0 = R_1 = 0$ to find that these equalities hold at $x_0 = 19.916$. There is a narrow range of initial stocks, $x_0 \in (19.897, 19.916)$ for which $R_0 > R_1 = 0$. For initial stocks in this range, period 1 rent is zero, but period 0 rent is (slightly) positive, because a larger stock reduces period-1 extraction costs.

The dotted line shows the graph of $y_0 = 0.5x_0$, where period-0 extraction equals half the initial stock. Because the solid graph lies above the dotted line, the figure implies that period-0 extraction always exceeds period-1 sales; thus for this example, the price rises over time.

4.6 Summary

Extraction costs may depend on the remaining resource stock, and the marginal extraction costs might increase with the level of extraction. We introduced a parametric cost function that has these features. We used both the standard method and the perturbation method to obtain the necessary condition for optimality in a two-period nonrenewable resource problem.

If demand is low relative to extraction costs, it might be optimal not to exhaust the resource. We therefore have to consider both possibilities, that the resource is or is not exhausted. If the resource is exhausted, then the resource constraint means that extraction of an additional unit at $t = 0$ requires an offsetting reduction in extraction at $t = 1$. If the resource constraint is slack, it is not necessary to make this offsetting change at $t = 1$.

In competitive resource markets, rent is defined as *price minus marginal cost*. Under monopoly, rent is defined as *marginal revenue minus marginal cost*. Recognizing this difference in the definition of rent under a competitive firm and under a monopoly, we can express the equilibrium condition for both markets in the same manner:

> Rent in period 0 (R_0) equals the present value of the rent in period 1 (R_1) plus the cost increase due to a marginal reduction in period-1 stock:
>
> $$R_0 = \rho \left(R_1 - \frac{\partial c\left(x_1, y_1\right)}{\partial x_1} \right).$$

The firm never extracts where rent is negative; rent is either strictly positive or it is zero. We can use this fact, together with the equilibrium condition, to solve for the equilibrium, given specific functional forms and parameter values for costs and demand. To do this, we first solve the problem under the *assumption* that the resource is exhausted in two periods. If this solution implies $y_0 > x_0$ (so that $x_1 = x_0 - y_0 < 0$), our assumption is incorrect (because it violates a nonnegativity constraint). In that case, we know that all of the resource is extracted in period 0 ($y_0 = x_0$). If the solution satisfies the nonnegativity constraint, we then determine whether it satisfies the condition $R_1 \geq 0$. If "yes", then we have the correct solution. If "no", then we know that the resource is not exhausted; in this case, the condition $R_1 = 0$ provides the third equation needed to solve the model.

4.7 Terms, study questions, and exercises

Terms and concepts

Binding constraint, slack constraint, stock-dependent and stock-independent costs, perturbation, rent, quasi-rent.

Study questions

1. Given an inverse demand function $p(y)$, an extraction cost function $c\left(x, y\right)$, a discount factor ρ, and an initial stock x_0: (i) Write down

the competitive firm's objective (= payoff) and constraints for the two-period problem. (ii) What assumption does this firm make regarding price in the two periods? (iii) Write down and interpret the optimality condition (= first order condition) for the firm. (Explain what the various terms in the equation mean.) (iv) Write down the definition of rent, and then restate the optimality condition in terms of rent in the two periods.

2. (i) In this two-period problem, what does it mean to say that the resource constraint is not binding? (ii) If the resource constraint is not binding, what is the value of period-1 rent? (iii) If the resource constraint is not binding, what is the value of period-0 rent? (iv) What does your answer to part (iii) tell you about the components of period-0 rent? (In answering parts iii and iv of this question you need to discuss the two situations where extraction costs are independent of, or depend on, the stock.)

3. Using the objective (= payoff) for the firm in question 1, and the assumption that the resource constraint is binding, describe how you can derive the optimality condition, first by eliminating the constraint, and second by the perturbation method. It is not necessary to take derivatives or do any calculation; just describe the steps.

4. Using the information provided in Chapter 2.4 and the competitive optimality condition, equation 4.5, write down and interpret the monopoly's optimality condition.

Exercises

1. This chapter considers only the competitive equilibrium. (a) For a general inverse demand function, and the parametric cost function, write down the monopoly's optimization problem in the two-period setting. (b) Under the assumption that the monopoly exhausts the resource, write down the equilibrium condition for the monopoly. (Hint: Review Chapter 2.4, especially the last subsection. (c) Say in words ("interpret") this equilibrium condition.

2. Figure 4.3 graphs two cost functions of the form $C\left(\sigma+x\right)^{-\alpha}y^{1+\beta}$, with $C>0$ and $\sigma>0$; the graphs hold $x>0$ fixed. (a) What can you

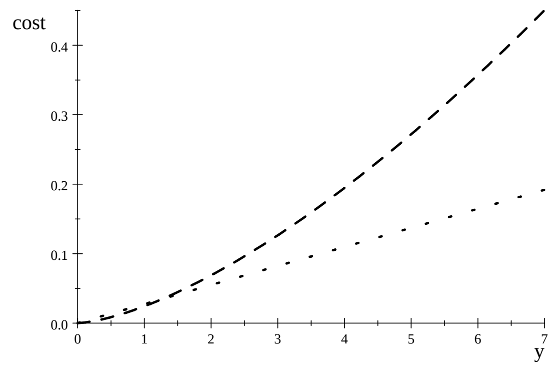

Figure 4.3: Graphs of two cost functions

conclude about β in these two functions? (b) What can you conclude from these graphs about the parameter α? Explain your answers.

3. (a) For the parametric cost function $c(x,y) = C(\sigma + x)^{-\alpha} y^{\beta+1}$, write the four partial derivatives:

$$\frac{\partial c(x,y)}{\partial C}, \frac{\partial c(x,y)}{\partial \sigma}, \frac{\partial c(x,y)}{\partial \alpha}, \frac{\partial c(x,y)}{\partial \beta}.$$

(Readers may want to consult Appendix B to review a particular rule of derivatives.) (b) Say in words what each of these partial derivatives mean. (This is a one-liner.)

4. Replace the general cost function used in the first order condition equation 4.5 with the parametric example given in equation 4.1. Next, rewrite the equation, specializing by setting $\beta = 0$. Explain in words the meaning of this equation.

5. Section 4.2.2 claims that the firm never extracts to a level at which marginal profit is negative. Explain, in a way that a non-economist will understand, why this claim is true.

6. In our two-period setting, the gain function for a candidate at which the resource is exhausted is

$$g(\varepsilon; y_0, x_1, y_1) = p_0 \cdot (y_0 + \varepsilon) - c(x_0, y_0 + \varepsilon) + \rho\left[p_1 \cdot (y_1 - \varepsilon) - c(x_1 - \varepsilon, y_1 - \varepsilon)\right],$$

and the gain function for a candidate at which the resource is not exhausted ($y_1 < x_1$) is

$$g(\varepsilon; y_0, x_1, y_1) = p_0 \cdot (y_0 + \varepsilon) - c(x_0, y_0 + \varepsilon) + \rho [p_1 \cdot y_1 - c(x_1 - \varepsilon, y_1)].$$

Identify the difference between these two functions, and provide the economic explanation for this difference, using a couple of sentences.

7. (a) Explain why period-1 rent must be zero if it is not optimal to exhaust the resource (so that the constraint $x_1 \geq 0$ is not binding). (b) With stock-dependent resource costs, explain why period-0 rent is positive even if it is not optimal to exhaust the resource. (c) With stock-dependent extraction costs, when it is not optimal to exhaust the resource, does rent rise, fall, or remain constant over time? Give a one sentence explanation.

8. Consider a two-period problem. Demand in a period is $a - by$ and extraction costs are $c(y, x) = \frac{y}{10+x}$. A competitive firm has initial stock x_0 and the discount factor ρ. (a) Identify the endogenous variables. Assuming that it is optimal to extract all of the stock, write down the equations you would solve in order to obtain the values of these endogenous variables. (b) Use these equations to obtain a single equation giving y_0 as an implicit function of x_0. (c) Now suppose that it is optimal NOT to exhaust the stock. Write the single equation that gives y_0 as an implicit function of x_0 in this case. (Hint: What must be true if it is optimal not to exhaust the stock in period 1? The answer to this question gives you y_1 as an explicit function of x_0 and y_0. Use this function and your answer to part (b) to answer (c).) (d) Is there a finite stock size above which period 0 equilibrium extraction sets price equal to marginal cost? Explain your answer in one or two sentences. (Hint: what are the two potential sources of period 0 rent?)

9. Under the assumption that the monopoly exhausts the resource in this two-period setting, write down the equilibrium condition for the monopoly that faces inverse demand $p = 20 - y$.

10. (Rent and quasi-rent for agricultural land.) Suppose that there is a fixed stock of unimproved land, $L = 10$. The value of marginal product of this land per year is $20 - q$, where q is the amount of land that is rented. (a) What is the equilibrium annual rental rate for land? (b)

How much would someone with an annual discount rate of r be willing to pay for this land? Recall equation 2.8. (The price of land is the amount that someone pays to buy the land; the rent is the amount they pay to use it for a period of time, e.g. one year.) (c) Suppose that if the land is improved, its value of marginal product increases by 2, to $22 - q$. What is the equilibrium annual rental rate if all land is improved? (d) What is the equilibrium price of improved land? (e) Suppose that the cost of this improvement is a one-time expense of 10. Assuming that the improvement (like the land) lasts forever, what is the critical value of r at which the landowner is indifferent between leaving (all of) the land in its unimproved state, and improving it?

Sources

Pindyck (1978) develops a model of extraction costs linear in extraction and decreasing in stocks.

Livernois and Uhler (1987) develop the point raised in the "Caveat" in Chapter 4.1, showing empirically that extraction costs rise with aggregate stock; for individual wells, they find that costs fall with the remaining stock.

Livernois (1987) estimates the extraction cost function for oil, finding that marginal cost is constant in extraction, and that the cost functions for different wells cannot be aggregated into an industry cost function

Chermak and Patrick (1995) estimate cost functions for natural gas, finding that costs fall with remaining reserves, but marginal costs fall with extraction.

Ellis and Halvorsen (2002), and Stollery (1983) provide empirical estimates of extraction costs.

Chapter 5

The Hotelling model

Objectives

- Interpret and use the optimality condition for the T-period problem.

Information and skills

- Understand the relation between the two-period and the T-period problems, and between their optimality conditions.
- Write down the objective, the constraints, and the Euler equation for the competitive firm in the T-period problem.
- Understand the relation between rent (and price) in any two periods.
- Understand the meaning of the "shadow value" of a resource.
- Understand the meaning of a Lagrange multiplier in a constrained optimization problem.
- Show that firms exhaust cheaper deposits before beginning to extract from more expensive deposits.

The intuition developed in the two-period setting survives when the resource can be used an arbitrary number of periods, $T \geq 1$. The perturbation method produces the optimality condition, known as the Euler equation in general settings, and the Hotelling rule in the resource setting. The definition of rent leads to a concise statement of this rule. Rent can be interpreted as the "shadow value" of the resource, the amount that a competitive resource owner would pay for one more unit of the resource in the ground. It can also be interpreted as the opportunity cost of extracting the resource.

After discussing the basics of the model, we consider distinct issues. We use an example to show that (in some circumstances) it is optimal to exhaust mines with low extraction costs, before beginning to extract from more expensive mines. We also discuss the parallel between the Hotelling rule and an asset pricing equation used in investment models. Finally, we obtain the necessary conditions for the monopoly by replacing "price" with "marginal revenue" in the equilibrium conditions.

For some resources (e.g. low cost Saudi oil) extraction will continue until the resource stock is physically exhausted. For other resources, stock-dependent extraction costs make it uneconomical to physically exhaust the stock. At some point, coal will become more expensive than other energy sources; coal deposits are unlikely ever to be exhausted. (Keeping the atmospheric temperature change below 2^oC, considered by some to be the maximum safe threshold, will require leaving over 50% of known fossil fuel stocks below the ground.) In general, there are two sources of rent (in our models): limited supply and stock-dependent costs. However, if it is optimal to stop extracting while some stock remains in the ground, the supply is not limited; in these cases, stock-dependent are the only source of rent.

Chapter 5.4 uses an example to illustrate the role of the non-negativity constraint on stocks. There we show how we to find the equilibrium value of T,and then find the equilibrium trajectory of extraction levels and corresponding prices. Elsewhere, we take T as given and we do not explicitly consider the non-negativity constraint on the stock.

5.1 The Euler equation (Hotelling rule)

Objectives and skills

- Write the competitive firm's objective and constraints.
- Write and interpret the necessary condition (the Euler equation).
- Understand how the perturbation method produces this condition.

We discuss the necessary condition for a fixed length of the program, $T + 1$, leaving the determination of T to Chapter 5.4. The competitive firm wants to maximize the present discounted sum of profits, subject to the resource constraint. The firm's optimization problem is:

$$
\begin{aligned}
&\max[(p_0 y_0 - c\,(x_0, y_0)) + \rho\,(p_1 y_1 - c\,(x_1, y_1)) + ... \\
&\rho^t\,(p_t y_t - c\,(x_t, y_t)) + ...\rho^T\,(p_T y_T - c\,(x_T, y_T))] \\
&= \max \textstyle\sum_{t=0}^{T} \rho^t\,(p_t y_t - c\,(x_t, y_t)) \qquad \text{subject to}
\end{aligned}
\tag{5.1}
$$

$x_{t+1} = x_t - y_t$, with x_0 given, $x_t \geq 0$ and $y_t \geq 0$ for all t.

The first order (necessary) condition for this problem is known as the Euler equation; in the nonrenewable resource setting, it is also called the Hotelling rule.[1] The equation is[2]

$$
p_t - \frac{\partial c\,(x_t, y_t)}{\partial y_t} = \rho \left[\underline{p_{t+1} - \frac{\partial c\,(x_{t+1}, y_{+1})}{\partial y_{t+1}}} \underbrace{- \frac{\partial c\,(x_{t+1}, y_{t+1})}{\partial x_{t+1}}} \right]. \tag{5.2}
$$

Equations 4.5 (for the two-period problem, where $T = 1$) and 5.2 (for general T), are identical, except for the time subscripts. Equation 5.2 must hold for all pairs of adjacent periods when extraction is positive: $t = 0, 1, 2...T - 1$.

Equations 4.5 and 5.2 have the same interpretation. If the firm sells one more unit in period t and makes an offsetting reduction of one unit in period $t + 1$, it receives a marginal gain in period t and incurs a marginal loss in period $t + 1$. The marginal gain in period t (the left side of equation 5.2) equals the increased profit, price minus marginal cost, due to the one unit increase in sales. The marginal loss in period $t + 1$ is the sum of the two terms on the right side of equation 5.2. The underlined term equals the reduced profit due to reduced sales, price minus marginal cost in period $t + 1$; the under-bracketed term equals increased cost due to the lower stock in period $t + 1$. Appendix E derives equation 5.2.

[1] The terms "Hotelling model" and the "Hotelling rule" are sometimes reserved for the case of constant marginal extraction costs, but we use the terms for general extraction costs.

[2] We sometimes show all subsbripts, as in the derivative $\frac{\partial c(x_t, y_t)}{\partial y_t}$. Where there is no possibility of ambiguity, to conserve notationwe sometimes drop subscripts, writing $\frac{\partial c(x_t, y_t)}{\partial y\backslash}$ or $\frac{\partial c(x_t, y)}{\partial y_t}$.

5.2 Rent and Hotelling

Skills and objectives

- Rewrite the Euler equation using the definition of rent.
- Explain the relation between rent in period t and in any other period, and understand how this relation depends on extraction cost.

In the competitive resource market, rent is defined as price minus marginal cost:

$$R_t = p_t - \frac{\partial c(x_t, y_t)}{\partial y_t}, \tag{5.3}$$

We will be interested in the *equilibrium value* of rent, it's value when the firm correctly solves its optimization problem. For brevity, we usually refer to this value merely as "rent". This (equilibrium), rent can be interpreted as the *opportunity cost* of extracting the resource: the loss from extracting the marginal unit now rather than at some other time. Current extraction reduces future profits, creating an opportunity cost. Rearranging the definition of rent, we have

$$p_t = \frac{\partial c(x_t, y_t)}{\partial y_t} + R_t.$$

This equation states that in a competitive equilibrium, price equals "full" marginal extraction cost, where "full" means the sum of the standard marginal cost and the opportunity cost (= rent). We use the definition of rent to write the Euler equation (Hotelling rule) more compactly, as

$$R_t = \rho\left[R_{t+1} - \frac{\partial c(x_{t+1}, y_{t+1})}{\partial x_{t+1}}\right]. \tag{5.4}$$

Constant marginal costs If marginal costs are constant (i.e. if $c(x, y) = Cy$), the Hotelling rule simplifies to

$$R_t = \rho R_{t+1} \;\text{ or }\; p_t - C = \rho(p_{t+1} - C). \tag{5.5}$$

The first equation states that the present value of rent is the same in any two adjacent periods with positive extraction. The constant cost model produces several important results.

- The present value of rent is the same in *any two periods where extraction is positive*:
$$R_t = \rho^j R_{t+j}. \tag{5.6}$$
The Hotelling rule states that the firm cannot increase its payoff by moving a unit of extraction between adjacent periods, t and $t+1$; this is a "no-intertemporal arbitrage condition". Equation 5.6 is more general: it states that the firm cannot increase its payoff by moving extraction between any two periods where extraction is positive (not merely between any two adjacent periods). The intuition for this relation is that a firm can sell the marginal unit in period t, invest the marginal profit (R_t) for j periods and obtain the return $(1+r)^j R_t$; alternatively, the firm can delay extraction of this marginal unit until period $t+j$, at which time it earns R_{t+j}. If there are no opportunities for intertemporal arbitrage, the firm must be indifferent between these two options, i.e.
$$(1+r)^j R_t = R_{t+j} \Rightarrow R_t = \frac{1}{(1+r)^j} R_{t+j} = \rho^j R_{t+j}. \tag{5.7}$$

- With constant marginal = average extraction cost, rent has a particularly simple interpretation. The value of the mine equals the initial rent times the initial stock:
$$\text{Value of mine } = \sum_{t=0}^{T} \rho^t (p_t - C) y_t = \sum_{t=0}^{T} \rho^t R_t y_t = R_0 \sum_{t=0}^{T} y_t = R_0 x_0. \tag{5.8}$$
The first equality is a definition: the value of the mine equals the present discounted stream of profits from extraction. The second equality uses the definition of rent, equation 5.3. The third equality uses equation 5.6 to replace $\rho^t R_t$ with R_0. The fourth equality uses the stock constraint: aggregate extraction ($\sum_{t=0}^{T} y_t$) equals the initial stock (x_0).

- We can determine the rate of change of price or rent by (i) multiplying both sides of the two equations 5.5 by $1+r$, (ii) using $\rho = (1+r)^{-1}$, and (iii) rearranging, to obtain
$$\frac{R_{t+1} - R_t}{R_t} = r \quad \text{or} \quad \frac{p_{t+1} - p_t}{p_t} = r - \frac{rC}{p_t}. \tag{5.9}$$

The first of these two equations says that *rent* rises at the rate of interest. The second says that *price* rises the rate of interest minus $\frac{rC}{p_t}$, i.e. price rises at less than the rate of interest if $C > 0$. For $C = 0$, the second equation simplifies to

$$\frac{p_{t+1} - p_t}{p_t} = r, \tag{5.10}$$

which states that in a competitive equilibrium where extraction is costless, price rises at the rate of interest: the competitive firm is indifferent between selling in any two periods if and only if the present value of the price is the same in the two periods. For $C > 0$, equilibrium price rises more slowly than the interest rate. In Chapter 3, we noted that constant extraction costs cause the firm to delay extraction, causing the initial price to be higher, and the later price to be lower than would have been the case for $C = 0$. Thus, $C > 0$ reduces $\frac{p_1 - p_0}{p_0}$ in the two-period setting. Equation 5.10 shows that in the T-period setting, a positive C lowers the rate of change of price at every point in time.

Stock dependent extraction costs With stock-dependent extraction costs, $(-\frac{\partial c(x_t, y_t)}{\partial x_t} > 0)$, the relation between (equilibrium) rent in periods t and $t + j$ is

$$R_t = \underline{\rho^j R_{t+j}} \underbrace{- \sum_{i=1}^{j} \rho^i \frac{\partial c\left(x_{t+i}, y_{t+i}\right)}{\partial x_{t+i}}}. \tag{5.11}$$

This equation shows that rent depends on two features, scarcity (the underlined term) and higher future extraction costs (the under-bracketed term). If the firm extracts an extra unit today and makes an offsetting reduction j periods in the future, the present value of the future loss in profit equals the underlined term. The under-bracketed term equals the present value of the higher pumping costs from periods $t + 1$ to $t + j$. The equilibrium value of current rent depends on rent and extraction costs in *future* periods. In equilibrium, rent is a "forward-looking variable", because it depends on prices and costs in the future.

The left sides of equations 5.3 and 5.11 are the same, but the right sides differ. This difference arises because the two equations have different meanings. Equation 5.3 is true merely because we decide it is true: it expresses a *definition*. (We are at liberty to define objects any way that we want, provided that we are internally consistent.) Equation 5.11, in contrast, is an *equilibrium result* that holds when the firm maximizes the present discount stream of profits.

5.3 Shadow prices and Lagrange multipliers

Objectives and skills

- Understand the meaning of the "shadow value" of a resource.
- Know that the perturbation method used above yields the same optimality condition as the Method of Lagrange.

How much would a resource owner be willing to pay to buy an additional unit of stock in the ground ("*in situ*")? The answer is known as the "shadow price" of the resource; the modifier "shadow" recognizes that the actual market for such a transaction could be hypothetical. *The shadow price at time t equals the equilibrium rent at that time, R_t.* This is a general relation, but it is particularly obvious in the case of constant marginal extraction costs. Here, equation 5.8 shows that the value of the mine is R_0x_0. The increase in this value, due to the increase in the stock, x_0, is the rent, R_0. The mine owner would be willing to pay R_0 for one more unit of the resource at time 0.

We used the perturbation method to obtain the equilibrium condition in the competitive resource market. The method of Lagrange provides an alternative. The firm's problem contains the T constraints $x_{t+1} = x_t - y_t$ for $t = 0, 1...T - 1$. To each of these constraints we assign a variable knows as the Lagrange multiplier. Having one more unit of the stock "relaxes" the constraint, i.e. makes it less severe. The Lagrange multiplier associated with a particular constraint equals the amount by which a "relaxation" in that constraint would increase the present discounted stream of profits. In the resource setting, the Lagrange multiplier associated with the constraint in a particular period equals the amount that the resource owner would pay for a marginal increase in the stock of the resource in that period; it equals the shadow price of the resource, which equals the rent.

A resource owner would be willing to pay exactly R_t (= the rent = the shadow value = the Lagrange multiplier) for an additional unit of the stock *in situ* at time t. This claim is not self-evident, because it might seem that an owner who receives one more unit of the resource would not simply extract that unit in the current period, and thus earn R_t. Instead, the owner could extract some of the extra unit in the current period, and some of it later. With that reasoning, it appears that the amount the owner would pay for the marginal unit might be greater than R_t. This conjecture is false because of the no-intertemporal-arbitrage condition: the owner has no desire to reallocate extraction over time. An owner who acquires one extra unit is indifferent between extracting it now, and earning R_t, or extracting it later. In either case, the *present value* of the owner's additional profit is R_t. The owner would therefore pay R_t for a marginal unit of the resource *in situ*.

5.4 Completing the solution (*)

Objectives and skills

- Understand the "transversality condition" and its role in solving for the equilibrium trajectories of price and extraction when T is either unconstrained or constrained.

If we are given a demand and cost function and the model parameters, we can use the Euler equation to solve for the equilibrium, much as in Chapter 4.3. If the owner is allowed to decide when to stop extracting, T is unconstrained. In this case, we have to solve for the optimal T along with the optimal trajectory of sales. An owner who is not able to extract beyond $\overline{T}$ (e.g. because a lease expires) faces the constraint $T \leq \overline{T}$. To explain the ideas as simply as possible, we restrict attention to the case of constant marginal extraction costs, C, and linear inverse demand function, $p = a - by$, with $a > C$. The parameter a is the choke price, the price above which demand equals zero.

5.4.1 T is unconstrained

For times $t < T$, where extraction is positive at both time t and at $t + 1$, it is possible to make a small *increase or decrease* in time t extraction, and make an offsetting change in the subsequent period $(t + 1)$. In contrast, at time T, current extraction is positive and extraction in the next period is 0: $y_T > 0 = y_{T+1}$. It is possible to make a small decrease in y_T and an offsetting increase in y_{T+1}, but (because negative extraction and a negative stock are infeasible) it is not possible to make a small increase in y_T and an offsetting decrease in y_{T+1}. Therefore, to test optimality of the candidate at time T, we need to consider only perturbations that decrease y_T. The optimality condition at time T, known as the transversality condition, is

$$[p_T - C] \geq \rho\,[a - C] \Rightarrow p_T \geq \rho\,[a - C] - C. \tag{5.12}$$

The second part of inequality 5.12 merely rearranges the first part, which has a straightforward interpretation. Under the candidate trajectory, period T is the last date at which extraction is positive. Therefore, under this candidate, the price in period $T+1$ is $a - b \times 0 = a$. A feasible perturbation reduces period T extraction, moving the marginal unit to period $T+1$. The cost to the firm of this perturbation is the marginal loss in profit at time T, the left side of (the first part of) inequality 5.12. The present value of the increased $T+1$ profits is $\rho\,[a - C]$. Inequality 5.12 states that the loss exceeds the gain. The firm does not want to delay extracting the final unit: it prefers to extract the final unit at time T. If this inequality did not hold, then T is not the optimal date to exhaust the mine. Thus, the transversality condition is a necessary condition for the candidate trajectory to maximize the mine owner's payoff.

Using equation 5.6, we also have

$$p_0 - C = \rho^t\,(p_t - C) \Rightarrow p_t = (1+r)^t\,(p_0 - C) + C. \tag{5.13}$$

Our goal is to find p_0. Once we know the value of this variable, the second part of equation 5.13 gives us the value of p_t. With this price, the demand function $y_t = \frac{a - p_t}{b}$ gives us period t extraction. How do we find p_0? For a given p_0, we use the second parts of equations 5.12 and 5.13 to write

$$\begin{gathered}(1+r)^T\,(p_0 - C) + C \geq \rho\,[a - C] + C \Rightarrow \\ T \geq -\frac{\ln\left(\rho\frac{a-C}{(p_0-C)}\right)}{\ln\rho}.\end{gathered} \tag{5.14}$$

Because sales are positive in period T, we also know that $p_T < a$. This inequality implies

$$\begin{gathered}a - C > (1+r)^T\,(p_0 - C) \Rightarrow \\ -\frac{\ln\left(\frac{a-C}{(p_0-C)}\right)}{\ln\rho} > T.\end{gathered} \tag{5.15}$$

Given p_0, T is the unique integer that satisfies equations 5.14 and 5.15; denote this integer as $T(p_0)$.[3]

Next, we use the stock constraint, which states that the sum of extraction, during periods when extraction is positive, must equal the initial stock:

$$\sum_{t=0}^{T(p_0)} \frac{a - \left((1+r)^t (p_0 - C) + C\right)}{b} = x_0. \tag{5.16}$$

Although we cannot find p_0 as a closed form expression of the model parameters, it is easy to solve equation 5.16 using numerical methods. One algorithm uses an initial guess of p_0 to evaluate the left side of this equation. If this calculation returns a value greater than x_0 we increase our guess of p_0, and if it returns a value less than x_0 we reduce the guess. Proceeding in this way, we improve the guess, until the left side is sufficiently close to x_0, giving an approximate solution. We use the approximation of p_0 to calculate T and then to calculate p and y in every period.

5.4.2 T is constrained

Here we consider the case where the owner is not able to extract beyond period T; we have the constraint $T \leq \overline{T}$. Denote the unconstrained value of T that we obtained above as T^{Endog} (endogenous T). If $T^{\text{endog}} \leq \overline{T}$, then the constraint is not binding, and the solution is as above. If, however, $T^{\text{endog}} > \overline{T}$, then the constraint is binding; here, the owner continues extracting until $\overline{T}$. In this case, we again find p_0 by solving equation 5.16, except that now instead of having the function $T(p_0)$ as the upper limit of the sum, we have the exogenous $\overline{T}$. If the solution to this equation is an initial price greater

[3] There is a unique integer that satisfies both of these inequalities because

$$-\frac{\ln\left(\frac{a-C}{(p_0-C)}\right)}{\ln\rho} + \frac{\ln\left(\rho\frac{a-C}{(p_0-C)}\right)}{\ln\rho} = 1.$$

In discrete time models, the terminal time must be an integer. For example, we might exhaust the resource in period $t = 18$ or $t = 19$, but we cannot exhaust it at $t = 18.3$. This "integer constraint" makes solving the discrete time model slightly more cumbersome than the continuous time mode, where we have no integer constraint. Therefore, many of the figures in subsequent chapters are constructed using the continuous time analog of the discrete time model presented in the text. Chapter 13.3 discusses the relation between the discrete and continuous time models.

than or equal to the extraction cost, C, we have the correct equilibrium. If the solution is an initial price less than C, we conclude that exhaustion of the resource is not an equilibrium: $\overline{T}$ is so small relative to the initial stock that exhaustion does not occur. In this case, the equilibrium price is C in every period, and rent is zero in every period.

5.5 The order of extraction of deposits

Objectives and skills

- Use the Euler equation to demonstrate that it is optimal to exhaust a cheaper deposit before beginning to use a more expensive deposit.

We have assumed that there is a single deposit, with extraction costs $c(x, y)$. We can think of this model as approximating a more realistic situation where there are many different deposits with different extraction costs. When marginal costs do not depend on either the stock of the level of extraction, a competitive equilibrium exhausts the cheaper deposits before beginning to extract from more expensive deposits.

Suppose, for example, that there are three different deposits, with stock size $x^a = 3$, $x^b = 7$, and $x^d = 2$, having associated constant average (= marginal) extraction costs $C^a = 4$, $C^b = 5.5$, $C^d = 7$. Figure 5.1 shows the average cost function for this example, as a function of remaining stock. The graph is a step function, because the extraction costs are constant while a particular deposit is being mined. Costs jump up once that deposit is exhausted, and it becomes necessary to begin mining a more expensive deposit. If instead of there being only three mines, with significantly different costs, there were many mines, with only small cost differences between the most similar mines, then the figure would approach a smooth curve, showing costs decreasing in remaining stock.

In order to show that, in a competitive equilibrium, cheaper deposits are exhausted before firms begin to extract from more expensive deposits, it is sufficient to consider the case where there are two mines, with constant extraction costs $C^a < C^b$. For exposition, we assume that one competitive firm owns the low-cost mine and another firm owns the high-cost mine.

The Euler equation must hold for both firms. In particular, for any two adjacent periods, t and $t + 1$, during which a firm is extracting, equation 5.9 must hold, with C replaced by C^a or C^b (depending on which firm is

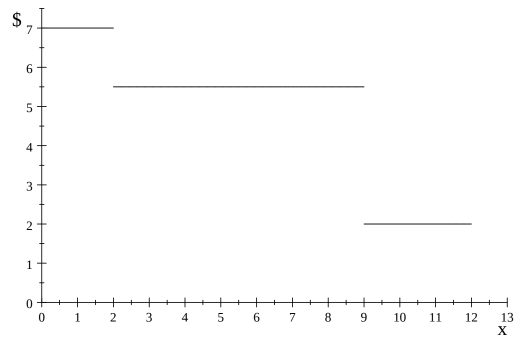

Figure 5.1: Extraction costs as a function of remaining stock, when there are three deposits, with individual stocks 3, 7 and 2, and corresponding constant marginal extraction costs, 4, 5.5., and 7

extracting). There might be a single period when extraction from both deposits occurs. For example, during the last period when the low cost firm extracts, its remaining stock may be insufficient to satisfy demand at the equilibrium price. In this case, there is a single period when both firms extract a positive amount. Extraction of high-cost deposits begins only after or during the last period when extraction from the low-cost deposit occurs.

To verify this claim, consider any two adjacent periods, t and $t+1$, during which the low-cost firm is extracting. We need to show that the high-cost firm does *not* want to extract in t, the first of these two periods. If the high-cost firm did want to extract in period t, then the claim would not be true, because in that case extraction of high-cost deposits occurs in a period prior to the period when the low-cost deposits are exhausted.

Because the low-cost firm extracts in both t and $t+1$, equation 5.9 implies

$$(p_t - C^a) = \rho\,(p_{t+1} - C^a)\,. \tag{5.17}$$

Some manipulations (Box 5.1) show that equation 5.17 implies

$$p_t - C^b < \rho\left(p_{t+1} - C^b\right). \tag{5.18}$$

This inequality implies that the high-cost firm strictly prefers to extract nothing in period t. If this firm were to extract a unit in period t, it would earn $p_t - C^b$. It could earn strictly higher present value profits by holding on to this unit and then selling it in period $t+1$. Therefore, it is not optimal for the high-cost firm to sell anything in period t.

Box 5.1 Derivation of inequality 5.18. Adding $C^a - C^b$ to both sides of equation 5.17 produces

$$\left(p_t - C^b\right) = \rho\left(p_{t+1} - C^a\right) + C^a - C^b.$$

Add and subtract ρC^b to the right side of this equation and then simplify the result to obtain

$$\left(p_t - C^b\right) = \rho\left(p_{t+1} - C^b\right) + (1-\rho)\left(C^a - C^b\right). \tag{5.19}$$

Because $1 - \rho = \frac{r}{1+r} > 0$ and $C^a - C^b < 0$, we have

$$(1-\rho)\left(C^a - C^b\right) < 0.$$

This inequality and equality 5.19 establish inequality 5.18.

Caveat The analysis above assumes constant marginal extraction cost in each mine. With more general cost functions, the relative costs of two mines might depend on the level of the stock or the rate of extraction in both, and it maybe efficient to extract from both simultaneously for many periods. For example, if mine i has costs $\frac{C_i}{2}y_i^2$ and mine j has costs $C_j y_j$, with $C_i < C_j$, marginal cost in mine i is lower than in mine j for low levels of extraction (where $C_j y_j < C_j$) but higher at high levels of extraction. If the Euler equations for both mines are satisfied at times t and $t+1$, then subtracting one equation from the other implies

$$y_{i,t} - \rho y_{i,t+1} = \rho\frac{C_j}{C_i}(1-\rho)$$

This equation determines the change over time in extraction from mine i, required for simultaneous extraction from both mines to be efficient.

5.6 Resources and asset prices

Objectives and skills

- Understand the relation between the basic asset pricing equation and the Hotelling rule.

Competitive equilibria eliminate opportunities for intertemporal arbitrage, both for natural resources and for other types of assets such as shares in companies. To illustrate that intertemporal arbitrage is important in many disparate contexts, we discuss the relation between the Hotelling rule and an asset pricing equation from financial economics.

By multiplying both sides of equation 5.4 by $\rho^{-1} = 1 + r$ and rearranging the result, we can write the Hotelling rule as

$$rR_t = R_{t+1} - R_t - \frac{\partial c\left(x_{t+1}, y_{t+1}\right)}{\partial x_{t+1}}. \tag{5.20}$$

Now consider the equilibrium price of an asset, such as shares in a company. There is no risk in our model, and a person can borrow a dollar for one year at the interest rate r. In this perfect information world, people know that next period, $t+1$, the price of the asset will be P_{t+1}, and they know that there will be a dividend on the stock of D_{t+1}. What is the equilibrium price of this asset today, in period t?

If the price of the stock at the beginning of period t is P_t, a person who can borrow at annual rate r can borrow P_t at the beginning of the period and buy a unit of the stock. They must repay $(1+r)\,P_t$ at the beginning of the next period. If they collect the dividend paid at the beginning of period $t+1$, and sell the stock, they collect $P_{t+1} + D_{t+1}$. Because the person has not used their own money to carry out this transaction (and thus incurred no opportunity cost), they must make 0 profits, which implies

$$\underbrace{P_{t+1} + D_{t+1}}_{\text{revenue}} - \underbrace{(1+r)\,P_t}_{\text{costs}} = 0.$$

Rearranging this equality gives:

$$rP_t = P_{t+1} - P_t + D_{t+1}. \tag{5.21}$$

Equation 5.21 is a no-arbitrage condition: it means that a person cannot earn profits by making riskless purchases and sales. The left side is the yearly cost of borrowing enough money to buy one unit of the stock. The right side is the sum of capital gains (the change in the price) and the dividend. To compare equations 5.20 and 5.21, recall that $-\frac{\partial c(y_{t+1}, x_{t+1})}{\partial x} \geq 0$, because a higher stock decreases or leaves unchanged extraction costs; this term equals the benefit that the resource owner obtains from lower future costs. It corresponds to the dividend, D_t in equation 5.21. The asset price, P, corresponds to R, the resource rent.

In the asset price equation, if the dividend is 0, then the price of the stock must rise at the rate of interest in equilibrium: the capital gain due to the change in the asset price must equal the opportunity cost of holding the stock, rP_t. A positive dividend makes investors willing to hold the stock at lower capital gains. The same reasoning explains why, in the resource setting, stock dependent extraction costs cause the equilibrium rent to increase at less than the rate of interest. The stock-dependent extraction costs play the same role in the resource setting as the dividend does in the asset price equation.

5.7 Monopoly

Objectives and skills

- Use previous results to write down and interpret the optimality condition for the monopoly.

We obtain the optimality condition under monopoly by using equation 5.2 and replacing price with marginal revenue, $MR_t = p_t\left(1 - 1/\eta\left(p_t\right)\right)$, where $\eta\left(p_t\right)$ is the elasticity of demand evaluated at price p_t. The Euler equation for the monopoly is

$$\begin{gathered} p_t\left(1 - \tfrac{1}{\eta(p_t)}\right) - \tfrac{\partial c(x_t, y_t)}{\partial y_t} = \\ \rho\left[p_{t+1}\left(1 - \tfrac{1}{\eta(p_{t+1})}\right) - \tfrac{\partial c(x_{t+1}, y_{t+1})}{\partial y_{t+1}} - \tfrac{\partial c(x_{t+1}, y_{t+1})}{\partial x_{t+1}}\right]. \end{gathered} \tag{5.22}$$

For the special case where extraction is costless and the elasticity of demand is constant (inverse demand is $p = y^{-\frac{1}{\eta}}$, and η is a constant), equation 5.22 simplifies to equation 5.10. In this special case, the monopoly and the competitive industry have the same price and sales path.

More generally, the monopoly and competitive outcomes differ. If extraction is costless but demand becomes more elastic with higher prices (as occurs for the linear demand function and many others) then monopoly price rises more slowly than the rate of interest. In this situation, the initial monopoly price exceeds the initial competitive price, so monopoly sales are initially lower than competitive sales. Here again, "the monopoly is the conservationist's friend".

Defining rent for the monopoly as marginal revenue minus marginal cost (instead of price minus marginal cost, as under competition), we can write the monopoly's optimality condition as in equation 5.4. The interpretation of this equation is the same as under competition, provided that we keep in mind that the definition of rent under monopoly differs from the definition under competition.

5.8 Summary

The perturbation method is as easy to use for the T-period problem as for the two-period problem. It leads to a necessary condition for optimality (equivalently, an equilibrium condition) that expresses current price and costs as a function of next-period price and costs. This equation is known as the Euler equation; in the resource setting, it is also known as the Hotelling rule. We can express this equilibrium condition in terms of rent.

The equilibrium market price and the rent in a period depend on future prices (or rent) and costs. Rent is a *forward-looking variable*. The Hotelling rule states that (equilibrium) rent in an arbitrary period equals the present value of rent in the next period, plus the cost increase due to a marginal reduction in stock. For the special case where extraction costs are independent of the stock, the Hotelling rule states that the present value of rent is equal in any two periods where extraction is positive. If extraction costs are zero, the Hotelling rule states that price rises at the rate of interest. For positive constant average extraction costs, price rises more slowly than the rate of interest. We showed how to solve the model numerically when the planning horizon, T, is either endogenous or exogenous.

The Hotelling rule has a close analog in investment theory, where the asset pricing equation states that the opportunity cost of buying an asset must equal the capital gains plus the dividend from owning the asset. By re-labelling the resource rent as the asset price, and the cost increase due to a lower stock as the dividend, the Hotelling rule becomes identical to this asset pricing equation.

We also discussed the following two points:

- Equilibrium rent in a period equals the amount that the resource owner would pay in that period for an extra unit of resource in the ground.

Rent equals the *shadow* price for this hypothetical transaction, which equals the Lagrange multiplier associated with the resource constraint in this period.

- It is optimal to exhaust cheaper deposits before beginning to use more expensive deposits. We confirmed this result for the case of mines with different constant extraction costs.

5.9 Terms, study questions, and exercises

Terms and concepts

Euler equation, Hotelling rule, transversality condition, asset price, capital gains, dividend, Lagrange multiplier, shadow value, *in situ*.

Study questions

1. (a) For a general inverse demand function $p(y)$ and the parametric cost function in equation 4.1, write down the competitive firm's objective and constraints. (b) Write down and interpret the Euler equation for this problem. (c) Without actually performing calculations, describe the steps of the perturbation method used to obtain this necessary condition.

2. (a) What is the definition of "rent" in the renewable resource problem for the competitive firm? (b) Use this definition to rewrite the Euler equation for the problem described in question #1. (c) Write down the relation between rent in period t and in period $t+j$ $(j \geq 1)$, and interpret this equation. In particular explain the difference between the case where extraction costs depend on the stock and where extraction cost does not depend on the stock.

Exercises

1. Write the Euler equation for a monopoly facing the demand function $p = 20 - 7y$ with the cost function $c(x, y)$ and discount factor ρ.

   Questions 2 and 4 require that you understand how an inductive proof works. Here is the context and the logic of this type of proof. You want to show that "something that depends on an integer j" is true for any positive integer j. In our context, the "something" is an equation. Inductive proofs use two steps. The first step shows that the "something" is true for $j = 1$. The second step shows that if the "something" is true for $j-1$, then it is also true for j. These two steps taken together mean that the "something" is true for $j = 1$; therefore it is true for $j = 2$; therefore it is true for $j = 3$...and so on.

2. (a) Use an inductive proof and the Hotelling rule (equation 5.4) to establish equation 5.6. (Note that this equation concerns the situation where extraction costs do not depend on stock.) In step 1, confirm (using equation 5.4) that equation 5.6 is true for $j = 1$. The second step requires that you show that *if* equation 5.6 is true for $j-1$, *then* it is also true for j. To accomplish this step, you *assume* that equation 5.6 is correct when you replace j by $j-1$. Using this assumption and equation 5.4, show that equation 5.6 must therefore be true for j. (b) provide a one- or two-sentence explanation of equation 5.6.

3. Fill in the missing algebraic steps that lead from equation 5.4 to 5.9 when extraction costs are constant.

4. (a) Use an inductive proof and the Hotelling rule (equation 5.4) to establish equation 5.11. (See question 2 above.) (b) Provide the economic explanation of equation 5.11 in a couple of sentences.

Sources

Hotelling (1931) is the classic paper in the economics of nonrenewable resources.

Solow (1974) and Gaudet (2007) provide time-lapse views of the role of the Hotelling model in resource economics.

Pindyck (1978 and 1980) made important contributions to analysis of the Hotelling model, including extensions to uncertainty.

Weitzman (1976) studied the optimal order of extraction from mines with different costs.

Dasgupta and Heal (1979), Fisher (1981) and Conrad (2010) provide graduate-level treatment of the nonrenewable resource model.

Berck and Helfand (2010), Hartwick and Olewiler (1986), and Tietenberg (2006), provide undergraduate-level treatment of this material.

Chapter 6

Empirics and Hotelling

Objectives

- Understand what it means to test the Hotelling model, and the practical difficulties of performing such a test.

Information and skills

- Summarize the main empirical implications of the Hotelling model.
- Have an overview of historical price patterns for several natural resources.
- Understand why data limitations complicate testing the Hotelling model.
- Understand why markets respond differently to anticipated versus unanticipated change, and the empirical implications of this difference..

Theories, in order to be useful, must generate hypotheses that can, at least in principle, be falsified (proven wrong). Models provide a means of stating a theory formally; like maps, they involve a trade-off between realism and tractability. A map of the world that consists of a circle is too abstract to be of any use. A map of the world containing all of the details of the world is equally useless.[1] Economic models help in identifying testable hypotheses and they can be useful in studying policy questions.

[1]Jorge Luis Borges' (very) short story "On Exactitude in Science" tells the tale of an empire in which cartography becomes so precise that the empire creates a map of its territory on a 1–1 scale. Later generations decide that this map is useless, except as a source of clothing for beggars.

The theory of the firm in neoclassical economics rests on the premise that rational firms attempt to maximize profits. ("Rational" does not mean "omniscient".) In a trade context, this theory means that firms care about the sum of profits across all markets, not merely profits in a single country. Exchange rates make it possible to express profits in different countries in a common unit of currency (e.g. dollars); transportation costs determine the importance of a commodity's physical location. In the natural resource context, the theory of the firm implies that firms care about profits in current and future periods, not merely in a single period. Discounting makes it possible to add up the profits in different periods, playing a role in the resource setting similar to exchange rates and transportation costs in the trade context.

The Hotelling model formalizes the theory of the firm in the natural resource setting. This model adopts the profit-maximizing premise and recognizes that resource stocks are finite, causing extraction costs to eventually rise, regardless of whether the resource is ever physically exhausted. This model has had little success in generating testable hypotheses, but can nevertheless be useful for policy analysis. The one hypothesis that is easily tested, based on constant extraction costs, is also easily rejected. Other hypotheses are testable in principle, but because of lack of data they cannot be tested directly.

A theory that generates no testable hypotheses cannot be corrected or rejected. At best, it provides a starting point for thinking about issues. Testing requires confronting a hypothesis with data, often using statistical methods. We avoid saying that a statistical test either accepts or rejects a hypothesis, and instead say that it either "fails to reject" or rejects a hypothesis. For example, a theory might imply that a particular elasticity is equal to 1. Using data and econometric techniques, we might obtain a 95% confidence interval of $(0.98, 1.12)$ for the elasticity estimate. Because the hypothesized value lies in this confidence interval, we would (for the 95% level of confidence) fail to reject the hypothesis that the elasticity equals 1. But we would also fail to reject the hypothesis that the elasticity equals (for example) 1.05. Because the elasticity cannot equal two different numbers, we can only say that the test fails to reject our hypothesis, not that it accepts the hypothesis. If the confidence interval did not include our hypothesized value, we say that the test rejects the hypothesis. In that case, we question (and perhaps improve) the theory that generated this hypothesis, or we question the data and econometric assumptions used to generate the test result.

The validity of the Hotelling model affects how we should think about resource markets and evaluate policy changes. Resource-pessimists worry that we will run out of essential resources. Resource-optimists think that prices respond to impending shortages, causing markets to create alternatives. This optimism rests on the belief that market outcomes are determined by rational profit-maximizing agents, i.e. that the Hotelling model describes resource markets. If the theory is correct (and markets are competitive), then the First Fundamental Theorem of Welfare Economics (Chapter 2.6) implies that the competitive equilibrium is efficient. In this case, the fact that nonrenewable natural resources are finite does not, in itself, create a basis for government intervention. There may, of course, be *distinct* motivations for intervention, e.g. concerns about equity or market failures, just as arise in many other markets. If, however, nonrenewable resource markets are inconsistent with even sophisticated versions of the Hotelling model, then resource markets are unlikely to be efficient and prices would provide little warning of impending scarcity, thus undercutting the basis for resource-optimism. Therefore, it is worth knowing whether the theory underlying the Hotelling model is a useful description of resource markets.

Given the uncertain empirical foundation of the Hotelling model, why do resource economists rely so heavily on it for policy analysis? The most persuasive answer is that the assumption of rational profit-maximizing firms is as plausible for natural resources as for other types of capital, where empirical testing has been more persuasive. It would be absurd to take literally the deterministic Hotelling model studied above, in which agents perfectly forecast future prices. But it is unlikely that resource owners – or owners of other types of capital – ignore the future in deciding how to use their asset. Investors make mistakes, but investors who systematically make mistakes are likely to be culled from the herd.[2]

Economists can rarely conduct the type of experiments that laboratory scientists perform. We seldom have one group of economies that serve as the "control group" and another that serve as the "treated group".[3] There-

[2] The lack of a better alternative is another reason for using the Hotelling model. Alternatives matter. How can you tell if someone is an economist? Ask them how their husband/wife/partner is. An economist will answer "Compared to what?"

[3] To test the efficacy of a new drug, researchers compare the outcome of people who receive the drug (the treated group) with those who receive a placebo (the control group). If we could conduct experiments with economies, the economies subject to the policy of interest would be in the treated group, and those without the policy would be in the

fore, economists use mathematical models, an alternative to bonafide (and infeasible) experiments to assess the likely effect of policy. For applications involving fossil fuels, the Hotelling model is key. By comparing a model outcome in the absence of a policy, and second outcome under a particular policy, we have at least some basis for evaluating that policy. Chapter 8 illustrates this procedure. If the underlying model is fundamentally wrong, then this kind of experiment is not informative. Therefore, the validity of the Hotelling theory is important.

6.1 Hotelling and prices

Objectives and skills

- Have an overview of the time trajectories of prices for major resources.

The simplest version of the Hotelling model, with zero extraction costs, implies that price rises at the rate of interest. With constant average extraction cost, $C > 0$, prices rise more slowly than the rate of interest. Figure 6.1 shows the profiles of real prices (nominal prices adjusted for inflation) for nine commodities. The dotted lines show the time trends fitted to this price data. The points of discontinuity in the dotted lines capture abrupt changes in the price trajectory. For most of these commodities there are long periods during which the price falls, and at least one abrupt change the price trajectory.

We need only price data to test the Hotelling model under the assumption of constant costs. In light of Figure 6.1, it is not surprising that research finds that this version of the Hotelling model is not consistent with data. We therefore consider versions of the model with non-constant costs, and then we consider more fundamental changes to the model or the testing procedure.

6.2 Non-constant costs

Objectives and skills

- Understand why it is difficult to test the Hotelling model with non-constant marginal extraction costs.

---

control group. An increasing number of such experiments have been conducted during the last fifteen years, but not at the macro-economy scale.

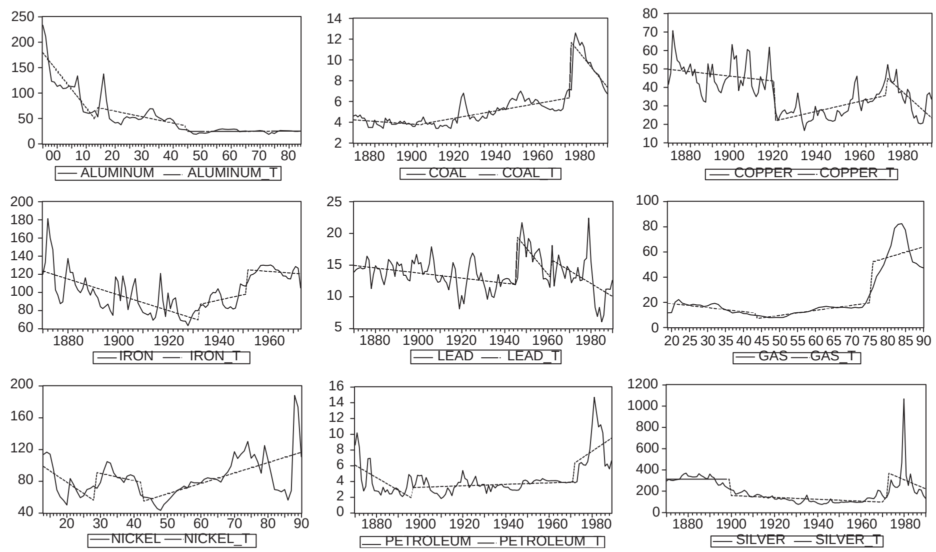

Figure 6.1: Real commodity prices. (Lee, List, and Strazicich, 2005)

Lack of data makes it difficult to test the Hotelling model with non-constant costs. The case where costs depend on extraction level, but not on remaining stocks, illustrates the problem. Here, the Hotelling rule is

$$p_t - \frac{dc\,(y_t)}{dy} = \rho \left(p_{t+1} - \frac{dc\,(y_{t+1})}{dy} \right). \tag{6.1}$$

The problem is that we do not observe marginal costs. If we assume (for example) that marginal costs equal $a + hy$, then we can write equation 6.1 as a function of prices and quantities and estimate the parameters ρ, a, h. If the parameter estimates are implausible, the researcher concludes that the model does not fit the data. This procedure involves a joint hypothesis that equation 6.1 is a reasonable description of behavior, and the marginal cost function $a + hy$ (or some alternative) is a reasonable description of the marginal cost. If the statistical tests reject our joint hypothesis, we do not know whether the rejection was due to the failure of one or both parts of the hypothesis.

If we had good data on costs (in addition to prices and quantities), then we could estimate a flexible cost function and have a reasonable degree of confidence in the resulting estimate of marginal cost. We would then be closer to testing the hypothesis involving behavior. Low quality cost data limits many fields of empirical economics, not just resource economics. A common procedure in other fields uses a firm's optimality condition, together with information about factor prices, to estimate marginal costs. Here, however, the empirical objective is to determine whether the optimality condition, the Hotelling rule, describes firms' behavior. It is not possible to both *assume* that the Hotelling rule holds and also to test whether it holds.

6.3 Testing extensions of the model

Objectives and skills

- Recognize that extensions of the model lead to a better fit with data, but increased difficulty of empirical testing.
- Understand the different equilibrium effects (e.g. on prices) of anticipated versus unanticipated changes, and the consequences for estimation.
- Understand the use of proxies in estimation.

The theory presented in Chapter 5 omits many real-world features, including: (i) the discovery of new stocks; (ii) changes in demand due to changing macro-economic conditions or the discovery of alternatives to the resource; (iii) changes in extraction costs due to changes in technology or regulation, (iv) and general uncertainty. We consider the empirical implications of these, and also explain consequences, for empirical testing, of the distinction between anticipated and unanticipated changes.

Owners of stock in a company have a claim on future profits of the company; owners of a resource stock have a claim on future profits from selling the resource. Asset prices in general, and resource prices in particular, are "forward-looking". Equilibrium asset prices (and resource rent) depend on expectations of future profits. Many firms (e.g. Amazon) had high stock valuations long before earning profits. The price of a company's stock depends on the market's perception of the company's *future* profitability. The equilibrium resource rent, and thus the current equilibrium resource price, depends on the resource firm's expectations of future prices and costs.

These expected future prices or events are "capitalized" into the asset price, meaning that the current asset price incorporates beliefs about them.

Asset prices change in response to surprises. For example, if the market has been expecting the Federal Reserve to increase interest rates, stock prices will not change much following the Fed's announcement of a rate increase, simply because the prices already take the likely rate increase into account. In contrast, if the market had expected the Fed to maintain low interest rates, the announcement of a rate increase comes as a surprise, and may have a significant effect on the market.

Changes in demand, technology (costs), resource stocks, and policy (e.g. taxes) all affect incentives to extract, and thus potentially change equilibrium prices. If resource owners care about future profits and are forward-looking, anticipated and unanticipated changes have different equilibrium effects; in addition, anticipated effects alter equilibrium outcomes even before they occur (Chapter 11.3.5). These features complicate the problem of estimating a model. In order to know whether an observed outcome is consistent with theory, it might be necessary to know (or estimate) the extent to which markets anticipated key events.

Stock changes We took the level of the resource stock in the initial period as fixed, and assumed that the stock falls over time, with extraction. In reality, firms invest in finding and developing new resource stocks, with uncertain success. Firms' attitude toward risk influence their decisions about extracting known reserves and about investing in the search to find undiscovered reserves. The theory presented above ignores the discovery and development of new reserves, and all risk.

Some extensions of the Hotelling model recognize that the exploration decision is endogenous. A simpler model takes the exploration decision as given, and treats the timing and the magnitude of new reserves as random variables. Rational firms factor this randomness into their extraction decisions. Large new discoveries create competition for previously existing deposits, lowering the value of those deposits, thus lowering resource rent. The large new discovery therefore causes rent on existing deposits to fall. Because the new stocks do not alter extraction costs of previously existing deposits, the fall in rent requires a fall in price. Thus, in a model with random discoveries, price rises between discoveries, but falls at the time of a large new discovery. In this scenario, the price path is saw-toothed, rising for a time and then falling at the times of new discoveries.

Royal Dutch Shell's experience with Arctic drilling illustrates the complexity of oil exploration. In 2005 Shell announced that it had previously overstated its proven reserves (known stocks) by more than 20%. This "surprise" (to the market) resulted in an overnight drop of 10% in the value of company. In an effort to increase proven reserves, Shell ramped up its Arctic exploration, buying drilling leases and drilling equipment, including a large rig named the Kulluck. In 2012 it received permission to begin exploration and towed the Kulluck into position. By the end of the year and a $6 billion investment, the Kulluck was destroyed in a storm, without having succeeded in drilling a well. Shell paused its exploration efforts in 2013; after receiving permits, it resumed exploration, but announced in September 2015 that it was abandoning these efforts. In August 2014, oil was $100/barrel, but it had fallen to $27/barrel by early 2016. "Unconventional" sources of oil, such as in the Arctic, require a price of about $70/barrel to be economical.

Changes in demand Unanticipated and long-lasting changes in demand can also change the price path. The Great Recession beginning in 2008 saw a reduction in aggregate demand, including a reduction in demand for many resources. Strong developing country growth from 1990 – 2010 increased resource demand. If a change in the economic environment causes firms to expect that future demand will be weaker than they had previously believed, then they revise downward their estimate of the value of their resource stock. This downward revision in rent requires a reduction in price, just as occurs following discovery of a large new deposit. Thus, over a period when firms are revising downward their projections of future demand, and thus revising downward their belief about the current value of a marginal unit of the stock, the equilibrium price rises more slowly than the (simple) theory predicts, and might even be falling. The deterministic (perfect information, no surprises) model ignores unanticipated changes in demand, although those random events are an important feature of the real world and could explain observed price falls.

Cost changes The simple Hotelling model ignores changes in costs, apart from those associated with changes in stock or extraction rates. In fact, the extraction cost function might shift up or down over time. Downward shifts are associated with declines in price, or with smaller price increases than the standard model predicts. Upward shifts lead to faster price rises than the model predicts. Technological advances lowered the cost of horizontal drilling and made hydraulic fracturing more effective, making it possible to develop previously inaccessible deposits. These cost reductions and the discovery of new deposits have similar equilibrium effects. Other changes, such as stricter environmental or labor rules, increase extraction costs. For example, in projecting extraction costs for new reserves, Royal Dutch Shell in 2014 included a (still nonexistent) carbon tax of \$40/Mt $C0_2$, or approximately \$17/barrel.

An example shows illustrates the effect of anticipated exogenously falling extraction costs. We replace the constant average cost C with

$$C(t) = C_0 + \frac{a}{1+ft}, \tag{6.2}$$

where C_0, a, and f are positive. As in equation 5.6, the no-arbitrage condition (the Euler equation) requires that the present value of rent is constant:

$$p_t - \left(C_0 + \frac{a}{1+ft}\right) = \rho\left[p_{t+1} - \left(C_0 + \frac{a}{1+f(t+1)}\right)\right]. \tag{6.3}$$

Figure 6.2 shows the graphs of price (solid), rent (dashed), and marginal cost (dotted) under the cost function in equation 6.2.[4] Falling costs put downward pressure on the equilibrium price, just as with a standard good. Discounting promotes an increasing price trajectory, as in the model with constant costs, C. Initially, the cost effect is more powerful, so the equilibrium price falls. However, the cost decreases diminish over time, and costs never fall below C_0. Eventually, the effect of discounting becomes more powerful, and the equilibrium price rises. Rent (price minus marginal cost) rises at the constant rate, r, and marginal costs steadily decline to C_0. Adding frequent small shocks and occasional large discoveries causes the graph in Figure 6.2 to become saw-toothed and bumpy, making it more closely resemble the graphs of actual time series shown in Figure 6.1.

General uncertainty A rich literature studies the role of uncertainty about new discoveries, changes in demand and extraction costs, and other features of nonrenewable resource markets. Uncertainty alters the expected

[4] Figure 6.2 corresponds to the *continuous time* version of the *discrete time,* with $a = 30$, $C_o = 10$, $f = 0.5$, and $r = 0.03$. The discrete time model leads to simpler derivations and more accessible intuition, but graphs in the continuous time model are easier to construct and interpret, simply because they are smooth, instead of step functions.

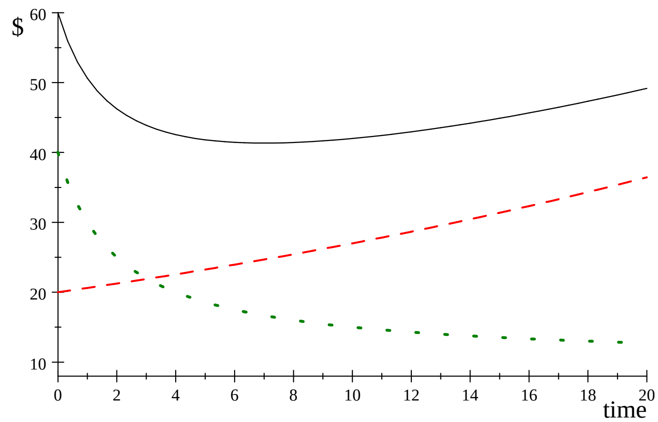

Figure 6.2: Equilibrium price (solid), rent (dashed) and marginal cost (dotted).

time path of rents, and thus of prices, even during periods when a large shock does not occur. Uncertainty complicates the already formidable problem of testing the Hotelling model, so researchers use proxies, observable variables closely related to the unobserved variable of interest: resource rent. In forestry the "stumpage price" is the price paid to a landowner for the right to harvest timber. Old-growth timber is a nonrenewable resource, and the stumpage price of this timber is a good proxy for rent. Researchers using this proxy find moderate support for the Hotelling model.

The stock price of a mining company is a proxy for the rent associated with the resource that the company mines. Both the observable stock price and the unobservable resource rent are forward-looking variables that reflect the expected value of the stream of future profits from owning one more unit of an asset. For rent, the asset consists of the resource. For the company, the asset is a composite of the resource, the mining equipment that the company owns, and many other factors, including intangibles such as reputation. These two assets are not exactly the same, so the company stock price is not a perfect measure of the resource rent. However, because the resource is a large part of the company assets, the two are closely related. The 2005 fall in Shell's stock price, mentioned above, following the downward revision in its proven reserves illustrates this relation.

Chapter 5.6 describes the change in asset prices under certainty. Under uncertainty, the Capital Asset Pricing Model (Box 6.1) shows that the equilibrium expected change in an asset price depends on the correlation between the price of the company and the price of a diversified portfolio of stocks (a "market portfolio"). If the two are negatively correlated, then the company tends to do well when the market does poorly, so owning stock in the company provides a hedge against market risk. (A hedge offsets the risk associated with a particular activity, e.g. investment in the market.) This hedge provides a benefit of owning the stock, and is analogous to the dividend discussed in Chapter 5.6: it lowers the expected capital gains needed to make investors willing to hold buy stock in the mining company./

> *Box 6.1 The Capital Asset Pricing Model (CAPM)* Investors can buy: (i) a risk-free asset (e.g. government bonds) that pays a safe return r_f; or (ii) a risky market portfolio with random return $\tilde{r}_m$ having expectation $E\tilde{r}_m = r_m$; or (iii) stock in a company that owns a nonrenewable resource, with a risky return of $\tilde{r}_c$ and an expected return r_c. The opportunity cost of investing in the stock market is the riskless rate r_f, and the "market premium" from investing in the stock market is $r_m - r_f > 0$. The "beta" of the mining company is
>
> $$\beta = \frac{\text{covariance}\,(\tilde{r}_m, \tilde{r}_c)}{\text{variance}\,(\tilde{r}_m)}.$$
>
> If all investors are rational and risk averse and have the same information and no borrowing or lending constraints, the CAPM shows that the equilibrium expected return to the mining asset must be $r_c = r_f + \beta\,(r_m - r_f)$. If the mining asset is negatively correlated with the market (its beta is negative), then investors are willing to purchase the mining asset at an expected return below the risk free rate, because the mining asset provides a hedge against market uncertainty.

The stock price of companies owning copper mines is negatively correlated with the return on a market portfolio. This negative correlation provides a hedge against market risk, thereby lowering the equilibrium rate of increase of the mining company's stock price. If the observed stock price is a good proxy for the unobserved rent, then evidence that one variable is consistent with theory provides some evidence that the other is as well. Using data on copper mining companies, this research "fails to reject" the Hotelling model.

In general, the market return might have positive, negative, or zero correlation with resource prices. A disruption in oil supply due to war and political turmoil, as in the 1970s, can increase oil prices and lower the return to the market, leading to negative correlation between oil prices and market returns. Strong economic growth can increase both oil prices and market returns, leading to positive correlation. Certain kinds of shocks are associated with positive correlation, and other kinds with negative correlation. Empirical evidence finds zero correlation between oil prices and the market return, but also finds that oil prices are negatively correlated with future economic growth. Large increases in oil prices preceded most of the post-World War II economic downturns.

6.4 Summary

Models are potentially useful for deriving testable hypotheses and for studying the effect of policy changes. The Hotelling model has not been successful in producing testable hypotheses; future chapters consider its role in analyzing policy. The Hotelling model can produce any number of hypotheses that can be tested *in principle*. The lack of reliable cost data makes it difficult to test most of these hypotheses.

With constant marginal and average extraction costs, the Hotelling model implies that price rises at less than the rate of interest. This statement about price increases is a joint hypothesis; it is based on a theory of firm behavior (profit maximization) and an assumption about costs. Statistical tests reject this joint hypothesis. Because the cost assumption is implausible, rejection of the hypothesis provides little information about the validity of the behavioral theory. For purposes of policy analysis, the behavioral theory is essential; assumptions about costs may be convenient, but not essential.

Adding real-world features, including non-constant marginal extraction costs, exploration, random shocks to stock, demand or cost, or systematic time-varying changes to costs, can generate simulated price paths that resemble observed price paths. Lack of data makes it difficult to directly test models that incorporate these kinds of realistic features. Because we cannot observe rents, they have to be constructed using estimates of marginal costs. Empirical tests of the Hotelling model using such constructed data thus involve a joint hypothesis, concerning the theory of profit maximization and assumptions about marginal costs. A few studies use proxies, such as stumpage fees for timber or the stock price of companies that mine resources, to indirectly test the Hotelling model. Those indirect tests provide modest support for the behavioral theory underlying the Hotelling model. The Hotelling model implies that the effect on resource prices of changes, e.g. in technology, policy, or demand, depends on the extent to which markets anticipated those changes. It is difficult to measure the degree of anticipation.

The Hotelling model has limited ability to generate testable hypotheses and it is not useful for predicting short or medium run changes in resource prices. The Hotelling model is nevertheless useful for studying the effect of policy changes, where we rely on the assumption of profit maximization, a hypothesis supported by other fields of economics.

6.5 Terms, study questions and exercises

Terms and concepts

Real versus nominal prices, joint hypothesis, control group, treated group, maintained hypothesis, anticipated versus unanticipated changes, proven reserves, proxies, hedge, market portfolio, market return, CAPM, stumpage, Occam's razor

Study questions

1. (a) Describe the simplest version of the Hotelling model's predictions about change in prices over time. (b) Describe important features of price series for actual nonrenewable resources, and contrast these to the predictions of the simplest Hotelling model. (c) What does it mean to empirically "test" the Hotelling model? Discuss some of the difficulties in conducting such a test. (Your answer should develop the idea that there are many versions of the Hotelling model, not just the simplest one. Your answer should also discuss the fact that a test of the Hotelling model is almost certainly a test of a "joint hypothesis", and explain why this matters.)

Exercises

1. Use the definition
$$p(y) - \frac{\partial c(x, y)}{\partial y} = R.$$

Suppose that extraction costs are convex in extraction ($\frac{\partial^2 c(x,y)}{\partial y^2} > 0$). (a) Give two examples of "news" about *future* events that would plausibly lead to a reduction in R. (b) Using economic logic (not math) explain briefly the effect of this type of news on the equilibrium values of y and p. (c) Use mathematics to confirm your answer to part (b). (Hint: recognize that y is an endogenous variable that depends on R; write $y = y(R)$ to reflect this dependence. Then take derivatives, with respect to R, of both sides of the definition of rent, to obtain an expression for the sign of $\frac{dy}{dR}$. Use the sign of this derivative to confirm your answer to part (b).

2. Extraction costs are $C(t)\,y$ with

$$C(t) = C_0 + \frac{a}{1+ft},$$

and $a > 0$ and $f > 0$. This cost function is non-stationary. (a) State in words what this cost function states in symbols. (b) Choose parameter values C_0, a, and f and graph $C(t)$. (c) Explain how to derive the Euler equation 6.3. (It is not necessary to perform calculations.)

3. Suppose that average extraction cost is constant, C, and the period t inverse demand function is

$$p_t = \frac{a}{a_0 + a_1 t} - y_t,$$

where a_0, a_1, and a are positive known parameters. (This type of model is called "non-stationary" because there is an exogenous time-dependent change – here, a change in the demand function.) (a) State in words what this demand function states in symbols. (b) Consider a competitive firm facing this demand function. Refer to Chapter 5 to write the Euler equation for this model (no derivation needed). (c) Explain what this equilibrium condition implies about the rate of change in prices. Provide an economic explanation. (d) This problem and the previous one show that nonstationary costs and nonstationary demand have different effects on the appearance of the Euler equation. Identify this difference, and provide an economic explanation for it. (e) The text discusses the effect, on the evolution of rent, of *unanticipated* changes in demand. Parts (b) and (c) of this question ask you to consider the effect of *anticipated* changes in demand. Summarize in a sentence of two the difference in the effect, on the evolution of rent, of unanticipated versus anticipated changes in demand. (f) Provide an economic explanation for this difference.

4. Using the parametric example in equation 4.1, discuss the equilibrium price effect of unanticipated cost reductions. To answer this question, you have to consider the various ways in which costs might fall; review your answer to Exercise 2 in Chapter 4.

5. Suppose that the return from investing in an index fund (a particular type of mutual fund) is a random variable $\tilde{r}_m$ with expectation $E\tilde{r}_m = r_m$ and variance σ_m^2; the return from investing in a copper mining company is a random variable $\tilde{r}_c$ and an expected return r_c and variance σ_c^2. Assume that $r_m > r_c$. The correlation coefficient between the two assets is $\rho < 0$. A person buys one unit of stock in the index fund and a units in the copper company. The return on this portfolio is the random variable $r_m + ar_c$. (a) What is the mean and variance of the return on this portfolio? [Hint: Use Google or consult a statistics textbook to find the formula for the mean and variance of the weighted sum of random variables, and also the formula that shows the relation between the correlation coefficient and the covariance of two random variables.] (b) Write the formula for a^c, the value of a that drives the variance on the portfolio to zero. (c) Would a risk-averse investor (one who dislikes risk) ever want to buy a portfolio that has more than a^c shares of the copper company for each share of the mutual fund? Explain in one sentence.

Sources

This chapter relies heavily on the surveys by Livernois (2009) and Slade and Thille (2009). Kronenberg (2008) and Krautkraemer (1998) also provide useful surveys.

Heal and Barrow (1980), Berck and Roberts (1996), Pindyck (1999), and Lee, List and Strazicich (2006) test the Hotelling model using price data.

Miller and Upton (1985) for oil and gas, Slade and Thille (1997) for copper, and Livernois, Thille, and Zhang (2006) for old-growth timber use proxies to test the Hotelling model.

Halvorsen and Smith (1991) use a restricted cost function to estimate shadow prices in the Canadian metal mining industry, rejecting the implications of the Hotelling model.

Chermak and Patrick (2001) use a model similar to Halvorsen and Smith, and fail to reject the Hotelling model. Chermak and Patrick (2002) subject the same data to four specifications, two of which reject and two of which fail to reject the Hotelling model.

Lin and Wagner (2007) find support for the Hotelling model for 8 of 14 minerals.

Malischek and Tole (2015), using uranium mining data, reject an extension of the Hotelling model that includes market power and exploration.

Pindyck (1980) pioneered the large literature on uncertainty and nonrenewable resources.

Farzin (1995) discusses the impact of technological change on resource scarcity.

Hamilton (2011) studies the relation between oil prices and economic activity.

Kilian (2009) discusses the distinction between oil demand and supply shocks.

Fama and French (2004) explain the Capital Asset Pricing Model and discuss its empirical significance.

Reuters (2014) reports Shell's assumption of carbon tax in projecting future extraction costs.

Funk (2014) describes Shell's experience with the Kulluck.

Chapter 7

Backstop technology

Objectives

- Apply the methods developed above to examine a backstop technology's effect on natural resource extraction.

Information and skills

- Know the meaning of and the empirical importance of backstop technologies.
- Understand why the existence of the backstop affects equilibrium extraction even before the backstop is used.
- Understand the effect of resource extraction costs on the timing of backstop use.

A "backstop" technology provides an alternative to the nonrenewable resource. Solar and wind power, and other methods of generating electricity, are backstop technologies for fossil fuels. We assume that the backstop is a perfect substitute for the resource, can be produced at a constant average cost, b, and can be supplied without limit. In contrast, the natural resource has a finite potential supply, given by the stock level. The backstop is available to the economy at large; any firm can use it.

The eventual use the backstop affects the equilibrium price and resource extraction paths, even during periods when the backstop is not actually used. The equilibrium level of resource extraction, and thus the resource price, in a period depend on firm's expectations of *future* prices. The backstop directly effect the market price in periods when it is used. Via expectations, the backstop also effects price and resource extraction even before it is used.

This insight is true generally. Anything that changes future resource prices – or, in an uncertain world, people's beliefs about these future prices – changes current extraction decisions. In the static model, competitive supply in a period depends on the price in that period. In the nonrenewable resource setting, however, current supply depends on *current and future* prices. The inter-relationship of markets across periods, in the resource setting, is analogous to the inter-relation of markets in the trade setting. In the trade example from Chapter 2, demand for tea in one country affects the equilibrium price and sales in the other country.

7.1 The backstop model

Objectives and skills

- Understand the simplest model of a backstop technology.

As above, we use y_t to denote resource use in period t. We need two new pieces of notation. Denote z_t as the amount of the alternative produced using the backstop technology in period t, and $w_t = y_t + z_t$ as supply of the resource plus the backstop good. In some settings, e.g. with fossil fuels and renewable power, it is natural to think of z as energy. Fossil fuels and renewable power are physically different, but can be expressed in common units of energy.

The assumption that the resource (e.g. oil) and the alternative produced using the backstop technology (e.g., solar power) are perfect substitutes means that the price in any period depends only on the sum of the resource and the backstop good brought to market, w_t. If both the resource and the backstop good are produced during a period, they must have the same price. This model of the backstop technology misses important real-world features. Just as coal is not a perfect substitute for oil, a low-carbon alternative such as solar power is not a perfect substitute for a fossil fuel. Solar power creates an "intermittency" problem: the power goes off when the sun goes down. Natural gas and coal do not suffer from this problem. Many machines that can run on fossil fuels have not been adapted to run on solar or wind power.

Taking into account the products' imperfect substitutability requires a more complicated model, having different prices for each product.

We assume a constant marginal cost for the backstop, but actual costs might either rise of fall with production levels or with cumulative production. Costs rise with output under decreasing returns to scale. Producing solar power uses land, and the most economical locations are likely to be used first. Therefore, subsequent solar farms may be more expensive than earlier units, leading to decreasing returns to scale. There may also be economies of scale, learning-by-doing, and technological advances that offset those cost increases. The unit cost of producing solar power in large scale solar farms might be lower than the unit costs of producing this power on houses and buildings, resulting in economies of scale. The history of new technologies shows that learning-by-doing and technological progress cause costs to fall over time. The cost of solar power is estimated to have fallen by a factor of 50 between 1976 and 2010. We overlook the products' imperfect substitutability and many of the complications associated with backstop costs in order to illuminate some of the significant features arising from the interaction of backstop and nonrenewable resource markets.

When society uses the backstop, price = marginal cost ($= b$), so

$$z > 0 \Rightarrow p = b. \tag{7.1}$$

The inverse demand function, $p(w)$, equals the price consumers pay for the quantity $w = y + z$.

Chapter 5 provides the tools needed to analyze the model of a backstop technology. The Euler equation establishes the relation between rent (prices minus marginal cost) in periods when extraction is positive. The transversality condition provides information about the final date of extraction, which we use to solve for the competitive equilibrium. We first use the constant marginal cost model, $c(x, y) = Cy$, to show how the presence of a backstop affects equilibrium resource extraction. We then consider more general extraction costs.

7.2 Constant extraction costs

Objectives and skills

- Identify and interpret the transversality condition in the backstop model with constant average extraction costs.

- Explain how and why the presence of the backstop changes the equilibrium price and extraction trajectories.

Here we assume constant average extraction costs, $c(x, y) = Cy$, with $C < b$ (otherwise, the resource is worthless) and b is less than the choke price (otherwise, the backstop is worthless). Apart from one minor difference, and different notation, this problem is the same as the problem discussed in Chapter 5.5, where mines have different extraction costs. There we denoted the extraction costs as C^a and C^b, and here we denote them as C and b. The minor difference is that here, the "expensive mine" is actually a backstop that, by assumption, can produce unlimited quantities at constant costs, b. In Chapter 5.5, the two mines have finite stocks.

Chapter 5.5 shows that it is never optimal to begin to extract from the more expensive mine while there remains available stock in the cheaper mine. Either the cheaper mine is exhausted before extraction begins from the more expensive mine, or there is a single transitional period during which extraction occurs from both mines. The same pattern occurs here, if we replace "the more expensive mine" with "the backstop technology". The trajectory consists of two intervals: only the resource is used during the first interval, and only the backstop is used during the second; there may be a single transitional period during which both are used.

Denote T as the last period during which resource extraction is positive. For $t < T$, resource extraction is positive in adjacent periods. The mine owner's objective and the optimality condition (Hotelling rule) are

$$\begin{gathered}\text{objective: } \max \textstyle\sum_{t=0}^{T} \rho^t \left[p_t y_t - C y_t\right], \\ \text{Hotelling: } p_t - C = \rho\left(p_{t+1} - C\right) \text{ for } t = 0, 1, 2..T-1.\end{gathered} \tag{7.2}$$

Now consider the mine owner's problem at time T, the last date at which (under the candidate solution) extraction is positive. If, in period T, the remaining resource stock is not sufficient to supply the entire market, then the backstop is also used in that period; in this case, the price in period T is b. If, in period T, the remaining resource stock leads to a price below b, then the resource supplies the entire market, and the backstop is not used until the next period, $T + 1$.

We consider both of these possibilities, using the firm's optimality condition at time T, the "transversality condition":

$$p_T - C \geq \rho\left[b - C\right]. \tag{7.3}$$

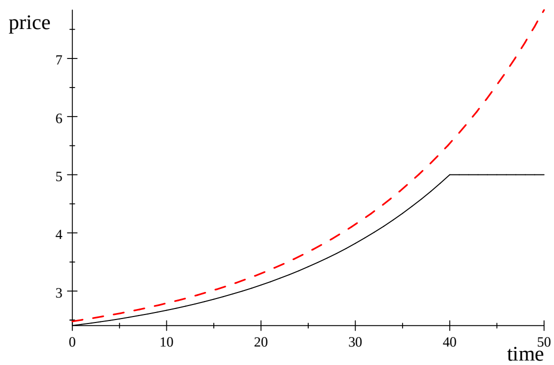

Figure 7.1: Solid curve shows the equilibrium price under the backstop. Dashed curve shows the equilibrium price without the backstop. Demand is $y = 10p^{-2}$, $C = 2$, $b = 5$, $x_0 = 42$, $r = 0.05$

In period $T+1$, when the backstop is being used, the price equals the backstop cost, b. The resource firm has the option, at time T, to reduce extraction by one unit, and to sell that unit in period $T + 1$. This perturbation reduces period T profits by $p_T - C$, and increases period $T+1$ profits by $b-C$. The present value, at time T, of the higher $T + 1$ profit is $\rho\,[b - C]$. Inequality 7.3 states that the profit reduction from lowering period-T sales is at least as great as the present value of the profit gain from increasing $T+1$ sales. If this inequality did not hold, then perturbing the candidate increases the present discounted sum of profits, implying that candidate is not an equilibrium. Inequality 7.3 is a necessary condition for optimality.

Given a demand function and parameter values, we can use the Euler equation and the transversality condition to find the equilibrium price and sales trajectories. Figures 7.1 and 7.2 graph these trajectories with the backstop (solid) and without the backstop (dashed).[1] These figures illustrate the general point that *the backstop affects the equilibrium sales and price path, even during periods when the backstop is not being used.* The future use of the backstop affects future prices, and these affect the current price, thus affecting current sales.

[1] Figures 7.1 and 7.2 show the graphs for the continuous time analog of the discrete time model. The continuous and the discrete time models are qualitatively similar, except that the transitional period, when the backstop and resource are used simultaneously, vanishes in the continuous time setting; there, the length of every period, including a transitional period, is infinitesimal.

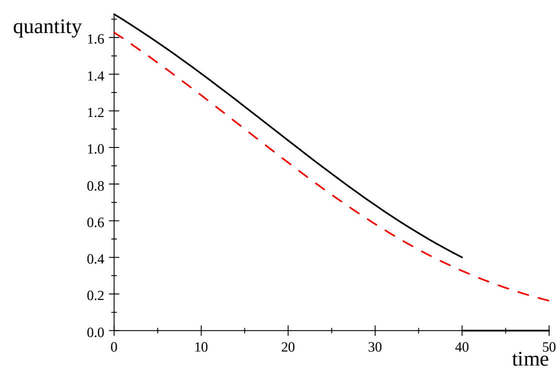

Figure 7.2: Solid curve: the equilibrium quantity of resource sales under the backstop. Dashed curve: the equilibrium quantity without the backstop.

Figure 7.1 is based on three facts: (i) During periods when the resource is being extracted, the no-intertemporal arbitrage condition (the Euler equation, or the Hotelling Rule) requires that price minus marginal cost rise at the rate of interest (the second line of equation 7.2). (ii) The resource is eventually exhausted, with or without the backstop (because marginal extraction costs is constant and less than the choke price). (iii) The price equals b when the backstop is being used (because the potential supply of the backstop is infinite, so price equals marginal cost in a competitive equilibrium).

Absent the backstop, we use Facts (i) and (ii) to determine the dashed curve in Figure 7.1. Given any value of the initial price, Fact (i) enables us to determine all subsequent prices, and thus to determine all subsequent sales levels. We identify the correct initial price using Fact (ii), requiring the sum, over time, of all of these sales levels to equal the initial stock. At times during which the price continues to rise, the resource has not yet been exhausted. In particular, at time $t = 35$ (when the vertical coordinate of the dashed graph in Figure 7.1 equals b), the resource has not yet been exhausted.

Now consider the solid curve, showing the price trajectory under the backstop. We might begin with the guess that the initial price, under the backstop, is the same as without the backstop. If that guess were correct, then the price (in the with-backstop scenario) follows the dashed curve until that curve reaches the backstop cost, b, at $t = 35$. We noted that at this time, the resource has not been exhausted (under the dashed price trajectory). Therefore, resource sales must continue after $t = 35$. By Fact (iii), in the with-backstop scenario, the price is constant at b after $t = 35$. Therefore, we conclude that during a period when the resource is being sold, the price is constant. However, this conclusion contradicts Fact (i) (the Hotelling Rule). Therefore, we know that our guess that the initial price under the backstop equals the initial price without the backstop, is incorrect.

Consequently, the initial price under the backstop must be either greater than or less than the initial price absent the backstop. If the initial backstop price were greater, then the price trajectory reaches the backstop cost, b, even before $t = 35$. Because the previous prices are higher than under our original guess, the previous sales are lower, so again the resource stock is positive when the price hits b, and again the resource is being sold while price is constant at b, again contradicting Fact (i).

Consequently, we know that the initial price under the backstop must be less than the price absent the backstop. The price reaches b at about $t = 40$ (instead of $t = 35$). During this longer period with lower prices, all of the resource stock has been sold, so that once the backstop begins to be used (and price remains at b) there is no more resource to sell; then, the Hotelling Rule does not apply.

7.3 More general cost functions

Objectives and skills

- Understand how the backstop changes the price and extraction trajectory when average extraction costs depend on either the level of extraction or on the resource stock.

The constant average (= marginal) cost function in the previous section is adequate for explaining the basic features of the backstop model, but it has two empirically false implications. (i) The model implies that, with the exception of a single transitional period, the resource and the backstop are never used in same period. However, when marginal costs increase with extraction rates, the resource and the backstop might be used simultaneously in many periods. (ii) The constant-cost model implies that the resource is physically exhausted before the backstop starts being used. However, when costs increase as the stock of resource falls, it may not be economical to physically exhaust the resource. We consider these two features separately.

7.3.1 Costs depend on extraction but not stock

Here we assume that marginal costs depend on the rate of extraction, but not the stock: $c = c(y)$, with $c'(y) > 0$ and $c''(y) > 0$; for example, $c(y) = Cy^{1+\beta}$. Both the resource firm and the "backstop firm" are price-takers. Due to this fact, we can think of the energy industry as consisting of a single representative firm that is able to use either or both the natural resource and the backstop. This firm's objective is:

$$\sum_{t=0}^{\infty} \rho^t \left[p_t (y_t + z_t) - c(y_t) - bz_t\right].$$

There may be many periods when both the resource and the backstop are used ($y > 0$ and $z > 0$). It is no longer true, in general, that society uses the backstop only when the resource is about to be, or has been, exhausted.

When marginal extraction costs increase with the extraction level, there can be either one or two distinct phases of resource extraction. In early periods, provided that the initial stock is sufficiently large, it is optimal to extract only the resource. During this phase, the Euler equation holds:

$$\begin{array}{c}\text{When the resource is being sold:}\\ p_t - c'(y_t) = \rho\left(p_{t+1} - c'(y_t)\right).\end{array} \tag{7.4}$$

Along this part of the trajectory, rent (price - marginal cost) rises to maintain a constant present value of rent. Rent can rise because price rises, or marginal cost falls, or a combination of the two.

Whenever the backstop is sold ($z > 0$) the condition that price equals marginal cost implies $p_t = b$. At some time, say T_1, firms begin to use the backstop, while continuing to extract the resource. During this phase, the price remains constant at $p = b$, but extraction falls (so c' falls). Price minus marginal cost rises, maintaining a constant present value of rent.

$$\begin{array}{c}\text{When both the resource and backstop are being sold:}\\ b - c'(y_t) = \rho\left[b - c'(y_{t+1})\right].\end{array} \tag{7.5}$$

At a later time, $T_2 \geq T_1$ the resource is exhausted and the backstop is the sole source of supply. The trajectory might consist of two phases of resource extraction (only the resource followed by both the resource and the backstop), or it might consist of a single phase (the resource and the backstop are used simultaneously).

Equation 7.5 is a no-arbitrage condition. Because the backstop is being used in both periods t and $t+1$, the price in both periods equals b. To interpret equation equation 7.5, consider a particular perturbation in which the firm extracts one more unit of the resource in period t, and makes an offsetting reduction in extraction in period $t+1$. This perturbation increases period t extraction cost by $c'(y_t)$. However, because the price remains at b, the perturbation does not alter total (resource + backstop) sales or revenue. Therefore, a unit increase in extraction requires a one unit reduction of backstop production in the same period. The reduction in period t total (backstop + extraction) costs due to this perturbation is therefore $b - c'(y_t)$, the left side of equation 7.5. The offsetting period $t+1$ decrease in extraction reduces costs by $c'(y_{t+1})$. However, because price remains constant, there must be a one unit increase in backstop sales, costing b. The net increase in future cost, associated with the perturbation, is therefore $b - c'(y_{t+1})$, and the present value of that additional cost equals the right side of equation 7.5. Along an optimal extraction path, the firm has no desire to reallocate the resource use: there are no opportunities for intertemporal arbitrage.

7.3.2 Stock-dependent costs

We now consider the case where the firm's extraction cost is $c(x, y) = C(\sigma + x)^{-\alpha} y$, i.e. average extraction cost depends on the resource stock, but not on the rate of extraction: costs are lower when the resource stock is higher. We assume that $b < C(\sigma + 0)^{-\alpha}$, which implies that it is not optimal to physically exhaust the resource. This assumption implies that there is a critical threshold of x, denoted $x_{\min}$, that solves $C(\sigma + x_{\min})^{-\alpha} = b$. It is never optimal to extract when the stock is below this level: there, the backstop is cheaper than the resource. If the initial stock of the resource is x_0, then the "economically viable" stock, i.e. the amount that will eventually be extracted, is approximately $x_0 - x_{\min}$.[2]

Here, the resource is not physically exhausted, but for low stocks it is

[2] Why "approximately" instead of "exactly"? We assumed that extraction costs depend on the stock at the *beginning* of the period. Thus, for example, if the stock is slightly above $x_{\min}$, it might be optimal to extract to a level slightly below $x_{\min}$. However, if the current stock is below $x_{\min}$, further extraction is uneconomical. This complication does not arise in a continous time model, where the economically viable stock is *exactly* $x_0 - x_{\min}$. Provided that the length of each period is reasonably small, say a year or so, the discrete time model and the continuous time model are similar.

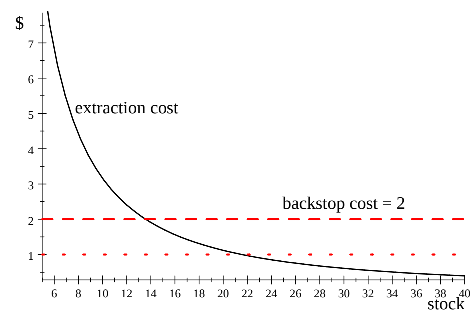

Figure 7.3: A lower backstop cost, b, increases the minimum economically viable stock.

not economically viable to extract more. Coal is one of the many examples of such a resource. Even apart from issues related to climate change, we will not use all of the coal on the planet, simply because at some point extraction costs exceed the cost of an alternative. Figure 7.3 shows a graph of stock-dependent average costs, the solid curve, and a backstop cost of 2 at the dashed line. In this example, extraction does not occur if the stock is below $x_{\min} \approx 13$. The cost of extracting stocks below this level exceeds the backstop cost. It is always optimal to drive the resource at least to the critical level $x_{\min}$. A trajectory that ceases extraction when the stock is above this critical level "leaves money on the table" (valuable resource in the ground). If the backstop cost falls from $b = 2$ to $b = 1$, the dashed line shifts to the dotted line, leading to an increase in the intersection, $x_{\min}$, and a decrease in the economically viable stock, $x_0 - x_{\min}$.

As with constant average extraction cost, there is an initial phase during which the backstop is not used, and an infinitely long phase during which only the backstop is used. There is at most a single period when both the resource and the backstop are used.

7.4 Summary

The backstop substitute affects the entire trajectory of the resource price, even before the backstop is used. This dependence reflects the fact that a firm's resource extraction decision is an investment problem. Extraction in a period depends on the relation between price in that period and in all subsequent periods. The backstop model drives home an important point:

the resource supply in a period depends on the entire trajectory of anticipated future prices. In the familiar static model, it is often reasonable to model the supply as a function current but not future prices.

Where average extraction cost is independent of the extraction level, there is at most a single transitional period during which both the resource and the backstop are used. In all other periods, only one energy source is used. If the marginal extraction cost increases with extraction, society might use both sources of energy simultaneously over many periods.

Variations of this model include the possibility that: the resource and the backstop are imperfect substitutes; the marginal backstop production costs rises with output (decreasing returns to scale) or falls with output (increasing returns to scale); and that the backstop cost falls with cumulative output (learning by doing). In these cases, the backstop and the resource might also be used simultaneously for many periods.

7.5 Terms, study questions and exercises

Terms and concepts

Backstop technology, transitional period, decreasing and increasing returns to scale, learning by doing.

Study questions

1. Explain why the presence of the backstop technology affects the competitive equilibrium even in periods before the backstop is actually used.

2. Different assumptions about extraction costs have different implications concerning the simultaneous use of the natural resource and the backstop. (a) Suppose that average (and marginal) extraction costs are constant. Describe the extraction profile of the resource, in relation to the production profile of the backstop. In particular, under what if any conditions are the two used in the same period? (b) Suppose that extraction costs are increasing in the rate of extraction; these costs do not depend on the remaining stock. Describe the extraction profile of the resource, in relation to the production profile of the backstop. In

particular, under what if any conditions are the two used in the same period? (c) Explain the source of the difference in parts (a) and (b).

3. Suppose that average (and marginal) extraction costs depend on the remaining resource stock, but not on the level of extraction. How does the magnitude of the backstop cost affect cumulative extraction?

Exercises

1. Use a proof by contradiction to establish equation 7.1. [*Hint:* To use this type of proof, write the hypothesis that states the "opposite" of what we want to prove, and derive a contradiction. Here, the hypothesis is "in some period when $z > 0$, $p \neq b$." This hypothesis is the "opposite" of equation 7.1, the statement that we want to verify. Then show that the hypothesis must be false, by considering individually the two possibilities, (i) "$z > 0$ and $p < b$" and then (ii) "$z > 0$ and $p > b$". The proof must explain why neither of these two statements can be true in a competitive equilibrium. Consequently, equation 7.1 must be true.]

2. In the constant cost model, explain why the resource is worthless if $C \geq b$. What is the resource rent in this case?

3. (a) In Chapter 7.3.1, why do we assume that $c'(0) < b$? (b) Explain why the assumption $c'(0) < b$ implies that it is optimal to exhaust the resource.

4. Chapter 7.3.1 claims that once firms begin using the backstop, they do not stop using it. Verify this claim, using a proof by contradiction. See the hint for problem 1 above.

5. Assume that average = marginal extraction costs are constant and independent of the remaining resource stock. Our model assumes that the backstop average = marginal cost is constant. Explain how the following modifications alter the equilibrium described in Chapter 7.2: (a) The backstop marginal cost is constant at a point in time, but falls exogenously over time (e.g. due to technological progress): the function $b(t)$, with $b'(t) < 0$ replaces the constant b. (b) the backstop marginal cost increases with the level of production; for example, *total*

backstop costs equal $\frac{1}{2}bz_t^2$, instead of bz_t as in the text. (z_t equals time t production of the backstop.)

Sources

Timilsina et al. (2011) review the evolution and the current status of solar power.

Heal (1974) is an early paper on natural resources and a backstop.

Dasgupta and Heal (1974) study the case where the backstop becomes available at an uncertain time.

Tsur and Zemel (2003) consider R&D investment that lowers the cost of the backstop.

Chapter 8

The Green Paradox

Objectives

- Use the Hotelling model to study the effects, on climate, of a policy that promotes a low-carbon fuel.

Information and skills

- Understand how a lower backstop cost affects cumulative extraction and/or the extraction profile.
- Explain why both of these changes might have climate-related consequences.
- Be able to synthesize this information to describe and then evaluate the "Green Paradox".

We discuss the "Green Paradox" for three reasons. First, the topic is intrinsically important because of its relevance to climate policy. Second, it provides an example of a situation where well-intentioned policies can backfire, a possibility that arises in many other contexts. Chapter 9 provides a more general perspective on this issue; the current chapter sets the stage for the general discussion by considering a specific example in detail. Third, the material shows how the Hotelling model can be used to illuminate a policy question. The Hotelling model makes the policy conclusions almost obvious, but without that model they would seem counter-intuitive. Formal models make it easier to understand real-world concerns.

Burning fossil fuels increases atmospheric stocks of greenhouse gasses. Scientific evidence shows that these higher stocks will affect the world's climate, possibly leading to serious environmental and economic consequences. The potential social costs of climate change include more serious epidemics, rising sea level, and increased frequency and severity of storms and droughts. Climate change may also lead to rapid and large-scale extinction of species, with unpredictable ecological, and ultimately social, consequences; temperature and precipitation changes might decrease agricultural productivity, worsening food insecurity; these changes may also induce massive human migration, worsening social conflict. Higher extraction, leading to higher atmospheric stocks of greenhouse gasses, worsens the climate problem. These risks have spurred interest in the development of "green" alternative energy sources, such as solar and wind power, which emit little or no carbon. Much of the political discussion concerns the use of public policy to reduce the cost of providing these alternatives.

We consider a particular policy, a subsidy that encourages firms to undertake research that decreases the cost, b, of a green backstop energy source. A cheaper backstop provides economic benefits during periods when it is used. The cheaper backstop also reduces the fossil fuel price trajectory before the backstop is used, benefitting energy consumers. However, if fossil fuel consumption was already socially excessive, the lower price of fossil fuels can lower aggregate welfare by further increasing carbon emissions. The possibility that an apparently beneficial change (the lower-cost backstop) harms society is known as the Green Paradox.

Other forms of this paradox build on the same general idea that resource owners *anticipate* a change that directly effects the market in the future. Possible changes include future carbon taxes or the future availability of substitutes to fossil fuels. Those future changes reduce future consumption of fossil fuels, benefitting the climate. But they induce changes in current behavior that might harm the climate. When these kinds of offsetting changes occur, their net effect on the climate may be ambiguous.

In an economy without market failures, cheaper energy increases social welfare. Although the lower backstop cost benefits society in this perfect world, the Fundamental Welfare Theorems imply that there is no need to subsidize green alternatives here. We do not provide public subsidies to competitive computer manufacturers, even though their innovations benefit consumers. In this perfect world, the market rewards the innovators by the amount needed to induce them to undertake the socially optimal level of innovation.

We are interested in public policy where market failures are important, e.g. due to the pollution externality associated with fossil fuels. Greenhouse gasses such as carbon dioxide are a classic example of a public bad. In the absence of "win-win opportunities" (Box 8.1) it is expensive for a country to reduce emissions. Those reductions lower future greenhouse gas stocks, reducing climate-related damages and creating global benefits. The individual country making the sacrifice to reduce emissions obtains only a small share of these global benefits. Due to this market failure, countries have inadequate incentive to reduce emissions, creating scope for global public policy.

Box 8.1 Win-win opportunities With "win-win" opportunities, the unilateral reduction of CO_2 emissions can benefit a country. Reducing $C0_2$ tends to also reduce local pollutants such as SO_2 and Total Suspended Particles (TSP), generating local health benefits. These kinds of "co-benefits" have been documented for China, and the Obama administration used these benefits as an additional justification for environmental rules announced in the summer of 2014. If the co-benefits are large relative to the cost of $C0_2$ reductions, the country can gain from the reductions even without taking into account the global benefit of lower atmospheric carbon stocks: a "win-win". Using sequestered carbon to improve soil or to reduce oil extraction costs create other win-win possibilities.

8.1 The approach

Objectives and skills

- Understand the distinction between the climate effect of changes in cumulative extraction and in the shape of the extraction profile.

Does a seemingly beneficial policy, such as a subsidy that lowers the cost of the backstop, actually help to correct the climate externality, or does it make things worse? The subsidy has two types of effects on the externality. First, the subsidy tends to reduce cumulative extraction of fossil fuels over the life of the resource, improving the climate problem. Second, the subsidy alters the extraction path, increasing extraction early on and decreasing extraction later. This "tilting" of the extraction path can worsen the climate problem. Either of these effects might dominate, so we cannot presume that the lower backstop price benefits the climate. We consider the two effects separately because a model that combines them is more complex, but not more insightful. Rather than speak generally of improvements in green technology, we use the backstop model from Chapter 7 with constant costs, b. An improvement in technology corresponds to a reduction in b, to $b' < b$.

The factors of production, e.g. scientists and lab space used to improve the technology, are costly. Proponents of green subsidies argue that the social benefits, arising from reduced climate-related damages, justify these investment costs. To focus on the Green Paradox, we ignore the investment costs. That is, we ask, "Even in the absence of investment costs, does society want the lower backstop costs?" A large literature discusses the merits of "Industrial Policy", governmental attempts to promote specific industries. All of the arguments for and against this type of government intervention also apply to green industrial policy.[1] The Green Paradox applies uniquely to green industrial policy, raising the possibility that society might not want the better technology even if it were free.

8.2 Cumulative extraction

Objectives and skills

- Recognize that with stock-dependent costs, lowering the backstop lowers cumulative extraction, making the Green Paradox less likely.

Chapter 7.3.2 considers extraction costs $= C(\sigma + x)^{-\alpha} y$, which increase as the remaining stock of the resource falls. There, zero extraction is optimal for stocks less than or equal to the threshold $x_{\min}$, the solution to $C(\sigma + x_{\min})^{-\alpha} = b$; cumulative extraction, over the life of the resource, is $x_0 - x_{\min}$. A lower backstop price, $b < b$, increases the threshold, lowering cumulative extraction (Figure 7.3).

By choice of units, we can set one unit of extraction equal to one unit of emissions, so reducing cumulative extraction creates an equal reduction of cumulative emissions. One short ton of subbituminous coal contains about

[1] An important criticism of industrial policy is that the government does a poor job of picking winners. Subsidies to Solyndra, a manufacturer of components to solar panels that went bankrupt in 2011, cost U.S. taxpayers $500 million.

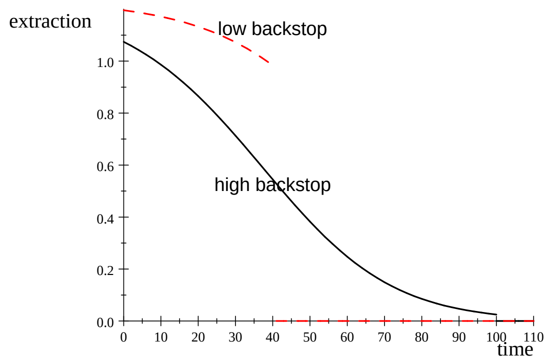

Figure 8.1: Extraction profiles under high and low backstop costs

3700 pounds of $C0_2$. Defining a "unit of coal" to equal a short ton and a "unit of $C0_2$" to equal 3700 pounds, one unit of coal equals one unit of $C0_2$.

With stock-dependent extraction costs, a lower backstop cost reduces *cumulative* emissions over the life of the resource. Insofar as climate-related damages arise because of cumulative emissions, the lower backstop cost reduces climate-related damages. Stock-dependent extraction costs therefore militate against the Green Paradox. If extraction costs are independent of the stock, the reduction in backstop costs has no effect on cumulative extraction and thus no effect on cumulative emissions.

8.3 Extraction profile

Objectives and skills

- Understand the effect of the backstop cost on the extraction profile.

A reduction in the backstop cost reduces future resource prices, thereby reducing the rent in earlier periods. This reduction in rent decreases the firm's opportunity cost of selling the resource. A reduction in the opportunity cost, like the reduction in any kind of (marginal) cost, increases equilibrium sales. Therefore, a reduction in the backstop costs leads to higher sales during periods that sales are positive. Because the resource stock is finite, it is not possible to increase sales at every point in time, so the lower backstop costs lead to earlier exhaustion of the resource.

Figure 8.1 shows extraction profiles under high and low backstop costs.[2] The important features of the figure are: (i) early in the program ($t < 41$ for this example) extraction is higher under the low backstop cost, and (ii) exhaustion occurs earlier under the low backstop costs. At $t = 41$, the resource is exhausted under the low backstop cost; exhaustion occurs at $t = 100$ under the high backstop cost. The arrival of an improved technology lowers the backstop cost (from b to $b' < b$). The lower backstop cost "tilts the extraction trajectory toward the present" (small t). Provided that the lower backstop cost does not make the resource stock worthless ($b' > C$), the arrival time of the new technology may be unimportant. In the example above, it does not matter whether the improved technology is available right away, at $t = 0$, or at $t = 40$; in either case, it is not used until $t = 41$.

Chapter 7.2 explains why a backstop lowers the price trajectory relative to the no-backstop case (before the backstop is actually used). Exactly the same reasoning implies that a reduction in the backstop cost, from b to $b' < b$, leads to a further reduction in the price trajectory. This reduction in the price trajectory corresponds to an increase in the sales trajectory. Given that there is a finite stock of the resource, the higher sales trajectory (during the period when extraction is positive) implies that the resource is exhausted sooner. Thus, a lower backstop cost implies that the sales trajectory is higher, during the period of positive sales, but the resource is exhausted sooner, as Figure 8.1 illustrates.

8.4 Why does the extraction profile matter?

Objectives and skills

- Understand three reasons why a tilt in extraction profile might increase climate-related damages.

[2] Figure 7.2 can be interpreted as comparing extraction profiles under an infinitely costly backstop and a backstop with a finite cost. Figure 8.1 compares extraction profiles under backstops with a high and a low cost. The two figures have the same message: lowering the cost of the backstop increases resource production before the backstop is used.

Figure 8.1 uses a continuous time model with constant average extraction costs, $C = 5$, $\rho = 0.95$ (a discount rate of 5%), demand $D = 10p^{-1.3}$, and an initial stock of $x_0 = 46$. The high backstop cost, $b = 100$, leads to the solid curve of extraction, and the low backstop cost, $b' = 6$, leads to the dashed extraction trajectory. In both cases it is optimal to exhaust the resource; rent is positive ($b > b' > C$). The continuous time curves are smooth, except for the points of discontinuity.

This section considers three reasons why a policy “tilts” the extraction profile to the present potentially worsens climate change: the tilt might make it more likely that we cross a threshold that triggers a catastrophe, such as rapid melting of the Antarctic ice sheet; the tilt might speed the rate of climate change, and society is worse off when change occurs more quickly; the tilt potentially creates a higher maximum stock level, and costs may be nonlinear in the stock. Figures 8.2 – 8.4 use the same assumptions as Figure 8.1.

8.4.1 Catastrophic changes

Figure 8.2 illustrates the possibility of crossing a threshold that triggers a catastrophe. The two extraction profiles in Figure 8.1 have different effects on the *stock* of atmospheric carbon. The distinction between stock and flow variables is critical. Here, emissions (equal to resource extraction, by choice of units) is a flow variable, measured in tons of carbon per year. The stock variable, the amount of atmospheric carbon, is measured in tons of carbon. The flow variable is measured in units of quantity *per unit of time*, whereas the stock variable is measured in units of quantity. Historical emissions, prior to the beginning of program ($t = 0$) determine the initial stock of atmospheric carbon. Some of the carbon entering the atmosphere migrates to other reservoirs, including the ocean and biomass. Although not literally correct, climate economists sometimes describe this migration using a constant decay rate for the stock. Fossil fuel emissions increase the stock of atmospheric carbon, and decay reduces the stock.

Because the initial stock level is historically determined, it is the same for both extraction paths at time $t = 0$. Later, for $t > 0$, the stock trajectory depends on the extraction (= emissions) trajectory. Because extraction is initially higher under the low-cost backstop, the stock grows more quickly in that scenario, relative to the high cost backstop scenario. Figure 8.2 shows the stock trajectories corresponding to the two extraction profiles taken from Figure 8.1. For approximately the first 70 years, the (dashed) stock trajectory under the extraction path corresponding to the low backstop cost lies above the (solid) trajectory corresponding to the high backstop cost.

Figure 8.2 also shows the flat dotted line at a stock of 38. If, for example, a stock above 38 triggers a catastrophe, then that catastrophe occurs in about 40 years under the low-backstop-cost trajectory, but never occurs under the high-backstop-cost trajectory. If the catastrophe is sufficiently severe, then

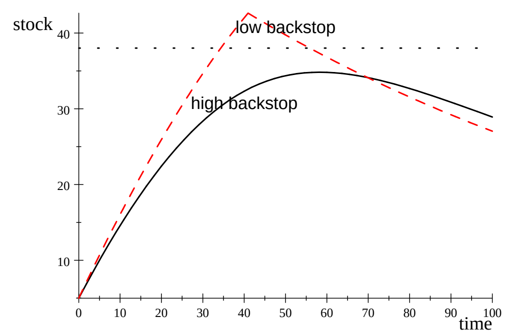

Figure 8.2: Atmospheric carbon stocks corresponding to the two extraction trajectories. Solid curve corresponds to high-cost backstop and dashed curve corresponds to low-cost backstop.

the future economic benefits arising from eventual availability of the low-cost backstop do not compensate society for the fact that this low-cost backstop "causes" the catastrophe. The lower cost backstop does not literally cause the catastrophe: the accumulation of stocks does that. But the lower cost backstop changes the competitive equilibrium extraction trajectory, thereby changing the stock trajectory, thereby triggering the catastrophe.

The model of the "carbon cycle" Figure 8.2 uses the assumption that the stock of carbon decays at a constant rate: the time derivative of the stock, $S(t)$, for this figure is

$$\frac{dS(t)}{dt} = y(t) - \delta S(t), \tag{8.1}$$

where $y(t)$ is emissions (= extraction) at time t and $\delta > 0$ is the constant decay rate. For this model, the stock rises ($\frac{dS(t)}{dt} > 0$) when $y(t) > \delta S(t)$ and the stock falls when this inequality is reversed.

This climate model is simple to work with, and therefore often used in policy models where the goal is to obtain insight, instead of making quantitative policy recommendations. The process that governs changes in atmospheric carbon (or more generally, greenhouse gas) stocks is much more complicated. In particular, a constant decay rate does not accurately describe the effect of emissions on the stock. In addition, GHG stocks likely cause damages indirectly, via the effect of the stocks on temperature or precipitation, instead of directly. The inertia in the climate system causes global average temperature and other climate variables to respond to changed GHG stocks with a delay. Therefore, the climate-impact of current emissions might increase over decades, before eventually diminishing.[3]

In addition to this delay, climate scientists have identified a number of positive feedback effects that might cause stocks to increase even if anthropogenic emissions were close to zero. For example, higher temperatures caused by higher stocks of greenhouse gasses might melt permafrost, releasing additional greenhouse gasses. Our model of catastrophes provides a simple way to think about this possibility. There may be a threshold level of atmospheric stocks that triggers such an event. However, the actual dynamics are much more complicated.

Figure 8.2 illustrates a possibility, but it does not establish that a particular outcome is likely. The figure does not show the unit of measurement of the stock variable, partly to defuse the danger that readers give it undue weight. Its key feature is that the *maximum* stock level under the low-cost backstop (the dashed trajectory) is above the maximum stock level under the high-cost backstop (the solid trajectory). Provided that the probability of catastrophe increases with the maximum stock level, this model (together with parameter assumptions) implies that the lower backstop cost increases the probability of catastrophe. This result is due to the fact that the initial emissions profile is higher under the low-cost backstop.

Box 8.2 The half life of the stock The half life of the stock equals the amount of time it takes half of a given stock to decay. With constant decay rate δ, $e^{-\delta t}$ of a unit emitted at time 0 remains at time t. Setting $e^{-\delta t} = 0.5$ and solving for t, produces the half life of the stock, $\frac{-\ln 0.5}{\delta}$. If the half life is between 100 and 200 years, and if we pick the unit of time to equal one year, then $0.0035 < \delta < 0.007$.

8.4.2 Rapid changes

Even in the absence of catastrophic events, tilting the extraction trajectory toward the present may harm society. If change occurs sufficiently slowly,

[3] Prominent climate-economics models, e.g. DICE, due to Nordhaus (2008), use climate components in which the major effect on temperature occurs five or six decades after the release of emissions. Recent evidence by Ricke and Caldeira (2014) suggests that the major effect occurs within the decade of emissions release.

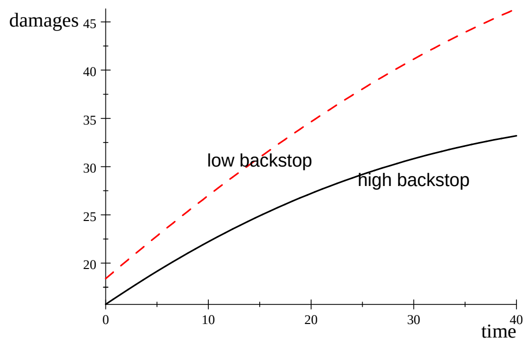

Figure 8.3: Damages are related to the stock and the speed of change in the stock. Solid curve corresponds to high-cost backstop and dashed curve corresponds to low-cost backstop.

society may be able to adapt to it with moderate costs. Over the very long run, society replaces most infrastructure. Climate-related change alters the speed at which this replacement must occur. If we know, for example, that rising sea levels will make some highways and bridges obsolete in 150 years, then we can divert investment away from maintaining these highways and bridges, and toward building more resilient substitutes. If we have to replace this infrastructure within the next 50 years, then we may be forced to write off much of the current infrastructure that would, absent rising sea levels, still be useful for decades.

As a simple way of modeling this dependence of climate-related costs on the speed of change, denote the stock at time t as $S(t)$ and the speed of change in the stock, the time derivative, as $\frac{dS(t)}{dt}$. Suppose that total damages depend linearly on the stock, and are convex increasing in the speed of change:

$$\text{Damages } = S(t) + 10\left(\frac{dS(t)}{dt}\right)^2.$$

With this formulation, marginal damages increase with the speed of change of the stock. From Figure 8.2, it is evident that the stock initially increases more rapidly in the dashed trajectory: its slope – the time derivative – is greater. The (historically determined) stock levels (under the two backstop costs) are exactly the same at the initial time. During the early part of the program, the stock levels are similar, so the damages related directly to the stock are also similar in the two scenarios. However, because the stock rises

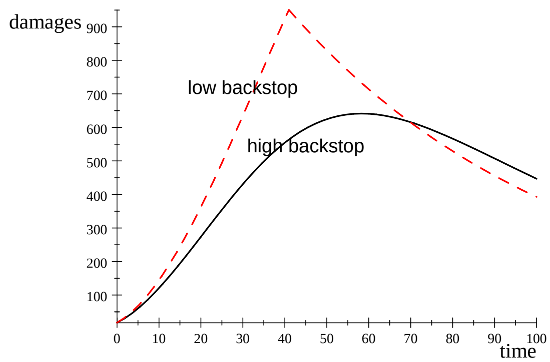

Figure 8.4: Damages are convex (quadratic) in the stock. Solid curve corresponds to high-cost backstop and dashed curve corresponds to low-cost backstop.

much more quickly in the low backstop cost scenario, the damages related to the speed of change of stocks is higher there. Therefore, early in the program, total damages are higher under low backstop costs. Figure 8.3 graphs of damages under the two backstops. Damages corresponding to the low-cost backstop (dashed curve) are higher early in the program.

8.4.3 Convex damages

In the previous example, damage is linear in the stock of atmospheric carbon: doubling $S(t) - S(0)$ doubles damage. If damages are convex in the stock, *marginal* damages are higher, the higher is the stock. The relation between atmospheric stocks and temperature change (e.g. feedbacks) or the relation between temperature change and damages, might create convex damages.

Figure 8.4 graphs convex damages equal to $S(t) + \frac{1}{2}S(t)^2$ under the high-cost (solid) and the low-cost (dashed) backstops. At the beginning of the program, damages in the two scenarios are the same, because the initial stock (determined by historical emissions) is the same. However, as Figure 8.2 shows, the stock becomes higher in the low-cost backstop scenario; with convex damages, the cost trajectory is higher in the low-cost backstop scenario.

8.5 Assessment of the Paradox

Objectives and skills

- Understand some of the nuances of the Green Paradox.

We explained the Green Paradox in the context of industrial policies that lower backstop costs. The same logic applies to carbon taxes that begin low and rise over time. Both of these policies lower future producer prices of fossil fuels, and therefore tend to lower current prices, increasing current extraction. Green subsidies and *future* carbon taxes are politically more palatable than policies that discourage current fossil fuel use. Because of their greater political appeal, and the resulting higher likelihood that they will be implemented, it is worth asking whether such policies have unintended consequences. Firms' current sales decisions, and thus the current equilibrium resource price, depend expectations of future prices. A lower expected future producer price decreases the scarcity rent, making it less attractive to store the resource rather than sell it today. Thus, lower expected future prices lower current price, increasing current consumption.

The Green Paradox exemplifies the constructive role that theory can play in informing policy, and also illustrates how easy it is to hijack theory to promote a particular agenda. Theory works best when it is simple enough to communicate easily. That simplification almost always requires focusing on a small set of issues to the exclusion of others. Once the theory has been understood in the simple setting, it is important to recognize its limitations.

8.5.1 Other investment decisions

The Green Paradox is usually studied in the Hotelling setting where fossil fuel extraction is the only investment decision. That treatment often ignores other investment decisions, including those related to the development of substitutes for fossil fuels or adaptations to anticipated policy, and those related to the discovery and development of new stocks of fossil fuels. When we recognize that businesses solve a host of investment problems, of which resource extraction is only one, the Paradox appears in a different light. Instead of providing a strong basis for rejecting green industrial policy, it merely reminds us that green industrial policy might have unintended consequences.

Changing the consumption portfolio

The immediate elimination of carbon emissions would be prohibitively expensive, but we are uncertain about the cost of moderate reductions. Resolving that uncertainty requires research and development in green alternatives, which likely require a combination of current R&D subsidies and future carbon taxes. Current subsidies reduce current investment costs, and the anticipation of future carbon taxes increases the expectation of the future profitability of current investment. Current carbon taxes increase the current profitability of low-carbon alternatives, but not their future profitability. Investment incentives depend on the anticipation of *future* profitability, because the fruits of current investment are available only in the future.

The Green Paradox focuses on the current response of resource *owners* to future changes in the market. However, resource *users* also have an incentive to adapt early to future changes. Consequently, anticipated future policy affects both demand as well as current supply. For example, the U.S. Acid Rain program was phased in over a decade, so coal producers and consumers were aware of future sulfur emissions constraints. This notification reduced the rent on high-sulfur coal, inducing owners to increase sales prior to the restrictions coming in to force; this is the supply effect examined by the Green Paradox. Power plants, the major consumers, recognized that the policy would make future emissions expensive. Businesses replace capital as it wears out; their replacement decisions depend on their expectation of future market conditions. The future implementation of the sulfur emissions policy gave power plants an incentive to replace aging capital stock with cleaner technology. Thus, the announcement of the future constraints increased current supply of dirty coal, but reduced near-term demand for that coal, leading to statistically insignificant effect on equilibrium consumption.

New sources of fossil fuels

The discovery and development of new deposits, such as tar sands deposits in Canada and oil off the coast of Brazil, involve substantial investment costs, including the costs of infrastructure needed to bring the oil to market. A fundamental rule of economic logic is to ignore sunk costs. The decision whether to develop the new deposits depends on the magnitude of the investment costs relative to potential profits. A green policy that lowers future expected resource prices, lowering future profits, can change the investment calculation. However, once firms have incurred the investment costs, the subsequent extraction profile does not depend on the now-sunk costs.

The Keystone pipeline would bring oil from the Canadian tar sands to U.S. refineries, and thence to world markets. Extracting and refining these deposits creates higher carbon emissions per unit of energy produced, compared to other petroleum deposits. Climate-change activists devoted substantial effort to influence U.S. policymakers to reject permits for this pipeline. Some of this opposition was due to concern about local environmental affects arising from possible leaks in the pipeline. Some of the opposition was for symbolic reasons, to show that the danger of climate change is great enough to justify derailing a project of national importance to Canada.

Delaying tactics can sometimes achieve a strategic goal. In 2012, with high oil prices, the pipeline looked like a solid business proposition. The economic viability of the project is less certain after the more than 50% drop in oil prices. Green industrial policy can increase uncertainty about the value of major new exploration and development efforts, delaying and possibly stopping these efforts. However, oil prices have historically been volatile (Figure 6.1) even before green industrial policy. Oil producers are accustomed to this volatility; the uncertainty about future green industrial policy (regulatory risk) is just one additional source of risk. As noted in Chapter 6.3, Shell already (in 2014) builds in a carbon tax to the cost of production, in anticipation that this tax will eventually be imposed.

Divestment from fossil fuel companies

The Green Paradox provides insight into possible effects of divestment from fossil fuel companies. Climate change activists encourage universities and pension and other investment funds to divest from fossil fuel companies, largely on the grounds of social responsibility. These activists draw parallels with the divestment from South Africa during the apartheid regime. By 2015, several universities (including Stanford) and cities (including Seattle, San Francisco and Portland) had begun to divest from coal companies. Proponents recognize that the divestment by a single fund, no matter how large, will have negligible effect on markets, but they hope that the publicity surrounding divestment debates will raise climate awareness.

There are economic, in addition to social-responsibility rationales for divestment. In 2015, Norway's parliament instructed the Government Pension Fund Global (GPFG), the world's largest sovereign fund, to divest from

114 companies, including 32 coal companies and several oil sands producers. (Ironically, shortly after the divestment decision, Norway's parliament voted to subsidize a national coal producer.) The economic rationale for the decision was that fossil fuel companies are overvalued (and therefore are poor investments) because the market does not account for the regulatory risk (e.g. future carbon taxes). This economic argument raises the question why Norway's parliament is better than the market at assessing a company's value.

If the divestment movement became powerful enough to lower the value of fossil fuel companies, it could have the perverse effect of increasing current emissions. The mechanism is the same as described in the Green Paradox. In the simplest Hotelling model with constant costs, we saw (equation 5.8) that the value of the firm equals the initial rent times the initial resource stock. In this setting, a decrease in the firm's value requires that the resource rent fall. The Green Paradox reminds us that a fall in resource rent – whatever its cause – increases current supply of fossil fuels. Possibly offsetting this effect, the fall in resource rent lowers incentives to find and develop new stocks, thus reducing cumulative supply. These two effects, greater current extraction but lower cumulative extraction, mirror the two effects studied in Chapters 8.4 and 8.2.

8.5.2 The importance of rent

The Paradox is relevant only for resources that have a substantial component of rent in their price. Chapter 6, notes that rent is a significant component of the price of oil, but a much smaller component of the price of coal. Resource-based commodities with low rent are similar to standard commodities. The Green Paradox has only slight relevance for such commodities, but so does the theory of nonrenewable resources.

The Green Paradox concerns policies that directly affect markets in the future, and only indirectly affect current markets. Some green policies have direct effects on current markets. For example, renewable fuel portfolio standards require a minimum fraction of energy produced using fossil fuel substitutes. Current solar and wind subsidies increase the demand for green energy sources today, not merely in the future. These sources are substitutes for fossil fuels, so the portfolio standards and the subsidies decrease the current consumption of fossil fuels, and do not lead to a Green Paradox.

8.5.3 The importance of elasticities

The significance of the Green Paradox depends on the price elasticities of supply and demand for fossil fuels. At least in the short run, demand is quite inelastic. With a low elasticity of demand, a lower current price transfers income from fossil fuel owners to fuel consumers, having modest effects on current consumption.

The Green Paradox is based on the assumption that a downward revision of beliefs about future energy prices would lead to a significant increase in current supply. Technical constraints may limit the supply response, at least in the short run. Many resource firms operate at or near capacity, and therefore have limited ability to quickly increase their supply. It may also be costly for them to shut down operations, lowering their flexibility to reduce supply. These considerations tend to reduce short run supply elasticities, reducing the significance of the Green Paradox.

8.5.4 Strategic behavior

The Paradox depends on the behavior of oil exporters, in particular OPEC, the oil cartel. OPEC is less powerful than a monopoly, because it faces a competitive fringe, but it is more sophisticated than the textbook monopoly because it understands that the demand function is *not exogenous*. The demand function in any period is *predetermined* by past events. Some past investments in infrastructure (e.g. highways) increase the current demand for fossil fuels, and other investments (e.g. development of the ethanol industry) decrease that demand. Because these investments have already occurred, the demand *function* (not the quantity demanded) at a point in time is predetermined.

OPEC observed that its oil embargo of the early 1970s changed behavior in importing countries, increasing conservation and the development of alternative supplies. OPEC wants to increase its rents, but it understands that the best way of doing that is not to extract every cent of consumer surplus available in the near term. OPEC's long-term strategy includes maintaining a reliable and reasonably priced source of petroleum, to discourage changes in behavior or the development of alternative sources that would reduce its future demand.

Green policies that reduce future demand can have ambiguous effects on current OPEC strategic behavior. One possibility is that the presence of these policies makes OPEC redouble its efforts to create a reliable and reasonably priced source of oil, in order to counteract the effect of the policies. The green policies encourage the development of green technologies, and OPEC may decide to try to offset that encouragement. How might OPEC go about achieving this goal?

- OPEC might think that the developers of the green technologies base their beliefs about future energy prices chiefly on current energy prices. With this view, OPEC could offset the subsidies to green technologies, discouraging their development, by reducing current price. In that case, OPEC's strategic response causes an even larger reduction in current prices, and thus a larger increase in current consumption than the standard Green Paradox suggests.

- Alternatively, OPEC might think that the developers of green technologies understand that future energy prices will depend on future stocks. With that view, OPEC could save more of its resource stock, in order to make credible its commitment to relatively low future prices. This commitment to low future prices requires a reduction in current extraction, thus working against the Green Paradox.

- A third possibility is that OPEC decides that efforts to discourage green substitutes for fossil fuel are doomed, thus diminishing its incentive to maintain stable and reasonable prices. If those efforts contributed to relatively low fossil fuel prices, then the reduction in those efforts increases current fossil fuel prices, working against the Green Paradox. That is, OPEC might decide that it is rational to exercise market power to the full extent possible, without worrying about the effect that high current prices have on the future demand function.

8.6 Summary

The Green Paradox illustrates the possibility that well-intentioned policies can backfire. The paradox potentially applies to policies that directly affect future energy markets, e.g. carbon taxes that begin in the future, or subsidy-induced reductions in the costs of backstop technologies that will be used in the future. These policies directly affect future demand for fossil fuels.

Because of the dynamic linkages in resource markets, those future prices affect resource owners' current supply decisions.

Policies that reduce *future* demand, makes it less attractive for resource owners to hold on to their stock, tilting the extraction trajectory toward the present, increasing current extraction and reducing extraction in some future periods. If extraction costs depend on the remaining resource stock, these policies also reduce cumulative extraction. Lower cumulative extraction, and the associated reduction in cumulative carbon emissions, benefit the climate. Tilting the extraction profile toward the present is likely to harm the climate. The higher earlier extraction likely increases the peak stock of atmospheric carbon, increasing the probability of a "catastrophe". The tilted extraction profile also increases the speed of change of atmospheric stock. Society may be worse off, the more rapid this change occurs. Finally, if marginal damages are increasing in the level of the stock, the tilted extraction profile is likely to increase damages. The net effect of policies that lower future demand for fossil fuels therefore depends on the balance between the effects of lower cumulative extraction and of higher earlier extraction.

The Green Paradox is valuable as a caution to policymakers, but practical considerations may limit its importance. The Green Paradox emphasizes the resource owners' incentives. Consideration of investment in resource exploration and development, and taking into account the externalities associated with investment in green alternatives, can overturn the paradoxical result.

8.7 Terms, study questions, and exercises

Terms and concepts

Research spillovers, business as usual, green industrial policy, green paradox, win-win opportunities, climate threshold, catastrophic change, stock and flow variables, decay rate, half-life of a stock, convex damages, predetermined versus exogenous.

Study questions

1. (a) State the meaning of the Green Paradox in the context considered in this chapter (where green industrial policy reduces the cost of a low fossil renewable alternative to fossil fuels). (b) Discuss some of the reasons that the paradox might occur in the case where fossil fuel

marginal extraction costs are constant. Your answer should include both a description of how, and an explanation of why, the industrial policy changes the extraction profile. It should also include a discussion of how and why this change in extraction profile might change climate-related damage. (c) Suppose now that extraction costs increase as the remaining resource stock falls. How and why does this different assumption about extraction costs affect the likelihood that the green paradox occurs?

2. Explain why the consideration of investment decisions other than the resource extraction decision might make the Green Paradox less likely.

Exercises

1. Suppose that a stock decays at a constant rate, δ, and that the "quarter-life" (defined as the amount of time it takes for 25% of an initial stock to decay) is 34 years. What is the numerical value of δ?

2. In Scenario A the damage caused by a stock S is fS (with $f > 0$ a constant). In Scenario B, the damage caused by a stock is FS^2 (with $F > 0$ a constant). (a) Graph damage and marginal damage in these two scenarios. (b) In which scenario are damages convex? (c) Explain in a sentence or two the meaning of convex damages.

3. Scenarios A and B are identical in every respect (e.g. demand function, initial resource stock, and extraction cost function), except for the following: in Scenario A, a backstop with cost b is available at time $t = 0$; in Scenario B it is known at time $t = 0$ that the backstop will not become available until $t = 49$. (a) Suppose that in Scenario A, the backstop begins to be used at time $t = 50$. What, if any, is the difference in extraction trajectories in the two scenarios? Explain your answer briefly. (b) In Scenario C, the time at which the backstop will become available is a random variable with expected value $t = 49$. Compare the equilibrium in Scenario C with those in Scenarios A and B and justify your conjectures. (Hint: You have all of the information needed for a complete answer to part (a), but not for part (b). The best you can do for part (b) is to make – and try to justify – intelligent conjectures.)

Sources

The DICE model due to Nordhaus (2008) is probably the most widely used model that "integrates" the economy and a climate cycle

Ricke and Caldeira (2014) provide evidence showing that major effect of emissions occurs within the first decade of emissions release.

Sinn (2008) (elaborated in Sinn 2012) is an early study of the Green Paradox.

Hoel (2008 and 2012) studies the role of extraction costs and demand characteristics.

Gerlagh (2011) distinguishes between a "weak" and "strong" paradox.

van der Ploeg and Withagen (2012) provide an in-depth analysis of the paradox.

van der Werf and Di Maria (2012) survey the literature.

Pittel, van der Ploeg and Withagen's (2014) edited volume brings together recent contributions.

Winter (2014) studies the Green Paradox in the presence of climate feedbacks.

Ellerman and Montero (1998) examine the effect of future emissions constraints on earlier sulfur emissions.

Di Maria et al. (2013) discuss the Green Paradox in the context of the U.S. Acid Rain Program.

Alberini et al. (2011) provide estimates of the elasticity of demand for electricity.

Vennemo et al. (2006) estimate the health benefits to China of reducing CO_2 emissions.

Lal (2004) discusses a win-win possibility involving carbon sequestration and soil enhancement.

Karp and Stevenson (2012) discuss green industrial policy.

Lemoine (2016) empirically tests the responsiveness of current coal price to expectations of future policy.

Schwartz (2015) describes Norway's divestment decision.

Chapter 9

Policy in a second best world

Objectives

- Understand the basics of designing policy under multiple market failures.

Information and skills

- Understand the Theory of the Second Best and the Principle of Targeting.
- Calculate and graphically illustrate the Pigouvian tax.
- Compare the optimal tax under monopoly and competition.
- Understand policy complements and substitutes.
- Understand how policies' interactions alters their welfare consequences.

Economists use the term "distortion" to mean any departure from an efficient allocation, or anything that causes such a departure. Examples include: (i) the gap between price and marginal cost arising from the exercise of market power; (ii) a gap between the private and social marginal production costs arising from a pollution externality; (iii) the gap (created by an income tax) between workers' incentive to supply labor (their after-tax wage) and firms' cost of labor. We usually emphasize efficient competitive markets, without distortions (Chapter 2.6), relegating market imperfections such as monopoly and externalities to a second tier of importance. We can make general statements about perfectly competitive markets, but not about markets with imperfections. To paraphrase Tolstoy: Perfect markets are all alike; every imperfect market is imperfect in its own way.

The focus on perfect markets yields valuable insights. For example, the theory of comparative advantage explains why trade potentially makes all participants better off, even if they have very different levels of development. The Hotelling model explains why scarcity *per se* is not a rationale for government intervention. For many markets, the perfectly competitive paradigm is also reasonably accurate. However, the emphasis on perfectly competitive markets sometimes seems like an elaborate justification of Dr. Pangloss' claim that "Everything is for the best in this best of all possible worlds." In fact, market failures are important, especially in natural resource settings.

The real world has multiple market failures, or distortions. We begin the study of these by introducing the "theory of the second best" (TOSB). A "first best" policy corrects a distortion or achieves an objective (e.g. raises government revenue) as efficiently as possible. It is difficult to rank all policies; we might not even know which to include. We may be able to say that a particular policy is not first best, but be unable to say whether it is 4'th or 17'th best. A policy is "second best" whenever it is not first best. The TOSB warns us against applying, in a second best world, the intuition obtained from the theory of perfect markets. *A policy intervention that seems likely to improve welfare might make matters worse.* In less extreme circumstances, a policy intervention might merely create unnecessary collateral damage. We also discuss a closely related idea: the Principle of Targeting. Examples help in developing intuition:

- Chapter 9.1 describes the TOSB and illustrates it using a trade example and the Green Paradox.

- Chapter 9.2 discusses the interaction of monopoly power and a pollution externality. The policy that corrects the externality under perfect competition is inappropriate under monopoly.

- Chapter 9.3 explains why political considerations often lead to inefficient policies.

- Chapter 9.4 compares pollution taxes and abatement subsidies, and explains the effect of extraneous distortions on optimal pollution policy.

- Chapter 9.5 considers the situation where an agricultural price subsidy magnifies the welfare cost arising from under-pricing a natural resource input.
- Chapter 9.6 shows that common sense may err in suggesting that one policy is an alternative (or substitute) for a second policy.

9.1 Second best policies and targeting

Objectives and skills

- Have an intuitive understanding of the Theory of the Second Best and the Principle of Targeting.

Different types of policies might alleviate a particular social, economic, or environmental problem. The choice of policy depends on political and social considerations. In democracies, and under most other types of governance, no single planner makes the policy decision. The benevolent "social planner" is a fiction, but one that provides a benchmark against which to compare the policies we observe.

In the partial equilibrium setting without market failures and without taxes, we take social welfare to be the present discounted stream of the sum of producer and consumer surplus. A competitive equilibrium maximizes this measure of welfare; it leads to the same outcome as the fictitious social planner. Here we are interested in market failures, so we need a broader definition of welfare. If, for example, the market failure arises from an unpriced externality such as pollution, we have to include the social cost of pollution and the fiscal cost or benefit of policies that attempt to remedy it. The first best policy maximizes social welfare; second best policies might increase social welfare, but they do so imperfectly, creating collateral damage or unnecessary costs.

Examples and the Principle of Targeting Some activists promote trade restrictions as ways of achieving environmental or resource objectives. Trade may increase environmentally destructive production, as occurred with shrimp harvesting that kills turtles. In the 1990s the U.S. imposed a trade restriction to redirect U.S. shrimp imports, hoping to decrease turtle mortality. A trade restriction might benefit the environment, but is seldom the optimal policy to achieve this goals. Turtle mortality was a consequence but not the goal of shrimp harvesting. The policy objective was to decrease turtle mortality, not to decrease trade. An efficient policy "targets" the environmental/resource objective. The U.S. trade restriction led to an international dispute that was resolved by the World Trade Organization (WTO). Although the WTO accepted that the U.S. had the right to use policies for the purpose of protecting international resources such as turtles, it also found that the U.S. policy contravened WTO law because it restricted trade unnecessarily. The dispute was resolved when the U.S. dropped its trade restriction but required exporting countries to use nets with "turtle excluding devices" that protected the turtles.

The Green Paradox provides another example of a policy that may be poorly targeted to an objective. The policy goal is to reduce carbon emissions. Low carbon alternatives to fossil fuels might help to achieve that goal, but green industrial policy that promotes these alternatives potentially changes the extraction profile in a way that harms the climate system. The net effect of green industrial policy might be positive, but is unlikely to be first best. First best policies, such as emissions taxes or cap and trade, directly target the environmental objective of reducing carbon emissions.

These examples illustrate the Principle of Targeting (POT). This principle states that a market failure (i.e., a "distortion" such as an unpriced externality), should be "targeted" as closely as possible. Many policies inflict collateral damage in correcting a distortion. The POT reminds us to be aware of this collateral damage or inefficiency, and to try to avoid it. In many cases, the application of the POT is straightforward. It is necessary to clearly identify the objective or the problem, and to distinguish between features that cause the problem and those that are associated with it. In the trade example, the problem is not trade, but that turtles are killed in catching shrimp. In the Green Paradox example, the problem is carbon emissions, not an excessively high backstop cost. The POT tells us that the efficient policy alters fishers' harvesting techniques in the first case, and reduces carbon emissions in the second.

9.2 Monopoly + pollution

Objectives and skills

- Use graphs and algebra to compare output under a social planner, a

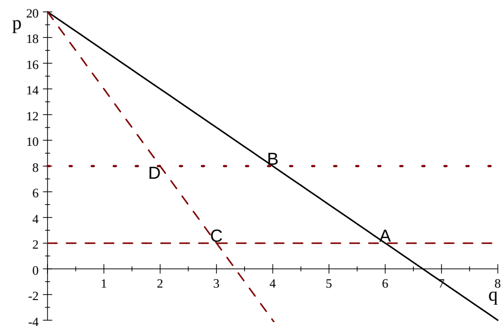

Figure 9.1: A: the competitive equilibrium with no pollution tax. B: the competitive equilibrium with the optimal (Pigouvian) tax $t = 6$. C: the monopoly equilibrium with no pollution tax. D: the monopoly equilibrium with the *non-optimal tax* $t = 6$

competitive firm, and a monopoly, in the presence of an externality.

- Show how a tax alters output under competition and monopoly.

An policy intended to alleviate one problem, might make another problem worse. A famous example of this possibility arises in a monopoly setting where production creates a negative externality, pollution. Figure 9.1 shows the inverse demand function $p = 20 - 3q$, the solid line, and the corresponding marginal revenue curve, $MR = 20 - 6q$, the dashed line. Average and marginal costs are constant at 2. Each unit of production (or consumption) creates \$6 of environmental damages; pollution is proportional to output, and social costs are proportional to pollution. The private cost of production here is 2 and the social cost of production, which includes environmental damages, is $2 + 6 = 8$.

An untaxed competitive industry produces where price equals marginal cost at point A. The monopoly sets marginal revenue equal to marginal cost, at point C. The symbol ν represents a unit tax; if $\nu < 0$, the policy is a subsidy: a negative tax is a subsidy. The optimal tax for the competitive industry, known as the Pigouvian tax, is $\nu = 6$. The socially optimal level of production and the price occur at point B. The tax $\nu = 6$ "supports" (or "induces") the socially optimal outcome: competitive firms facing this tax produce at the socially optimal level. The optimal tax causes firms to face the social cost of production, inducing them to "internalize" the cost of pollution. If the monopoly where charged the same tax, $\nu = 6$, it's tax-inclusive cost of production also equals the social cost. The monopoly facing $\nu = 6$ produces at point D. Absent the tax, the monopoly produces too little, relative to the socially optimal level: point C lies to the left of point B. The tax causes the monopoly to reduce output even more, lowering social welfare: the tax that is optimal in a competitive setting ($\nu = 6$) lowers social welfare if imposed on the monopoly.

This example illustrates the TOSB: a policy that improves matters in one circumstance might make things worse in another. In the competitive setting, there is a single distortion, arising from pollution. In the monopoly setting, there are two distortions, one arising from pollution and the second arising from the exercise of market power. A tax that fixes the first distortion makes the second one worse. For our example (but not in general), the net effect of the policy that is optimal under competition, lowers welfare under monopoly; the optimal policy under a monopoly is a subsidy (a negative tax). In general, the optimal pollution tax is lower under the monopoly than under competition, simply because the monopoly produces less than the competitive level.

The algebra The socially optimal level of production equates the marginal benefit of consumption (the market price) to the full social cost of production (the private cost plus the externality: $20 - 3q = 2 + 6$). The socially optimal level of production is $q^* = 4$. A competitive firm facing the tax ν produces where price equals private marginal cost plus the tax, $20 - 3q = 2 + \nu$, implying the production level $q^{\text{compet}} = 6 - \frac{1}{3}\nu$. The optimal tax for the competitive firm causes the competitive level of production to equal the socially optimal level ($q^{\text{compet}} = q^*$), which requires the tax $\nu = 6$, equal to the externality.

The monopoly facing a tax ν produces where marginal revenue equals production cost plus the tax, $20 - 6q = 2 + \nu$, implying $q^{\text{monop}} = 3 - \frac{1}{6}\nu$. The monopoly produces at the socially optimal level if $q^{\text{monop}} = q^*$, or $3 - \frac{1}{6}\nu = 4$, implying that $\nu = -6$, a subsidy.

A caveat Our example assumes: (i) a fixed relation between output and pollution, and (ii) constant social marginal cost of pollution. Assumption (i) means that the only way to reduce pollution is to reduce output. In reality, it is often possible to reduce pollution without reducing output, by using a more costly production method. Assumption (ii) also makes the message simple to deliver, but it can easily be dropped.

9.3 Collective action and lobbying

Objective and skills

- Use a payoff matrix to identify a noncooperative Nash equilibrium.
- Illustrate the collective action problem that arises with lobbying.

Understanding the collective action problem helps in making sense of political outcomes. A collective action is a costly action taken by a group, for the benefit of the group. People prefer other members of their group to incur the costs, while they share the benefits. Society may impose a solution to this problem by forcing group members to contribute, provided that a sufficiently large fraction of the group has voted to do so. U.S. marketing orders and union laws illustrate these kinds of imposed solutions. The U.S. Agricultural Marketing Agreement Act of 1937 obliges producers to participate in marketing orders; these might require minimal quality levels or limited production (in order to maintain high prices), or fees (to support generic advertising). About half of U.S. states have laws requiring workers to pay union dues to a legally recognized union, on the ground that all workers benefit from union representation in their workplace. The constitutionality of both marketing orders and of mandatory union dues has been challenged, with some success, in U.S. courts during the past quarter century. Plaintiffs object, for example, that they do not share the goals of the marketing order or the union, and that their enforced participation deprives them of their property or their right of self expression. What seems to one person a solution to the problem of collective action, appears to another as an infringement on liberty.

Real-world policies emerge from a political process, not as the dictate of a benevolent social planner. Political lobbying or naked corruption affects these outcomes. The Sunlight Foundation estimates that in the U.S. between 2007 – 2012, 200 companies spent \$5.8 billion in lobbying and campaign contributions, and received \$4.4 trillion in federal support or contracts: \$760 for each dollar contributed. Changes in U.S. law, notably the Supreme Court ruling in "Citizens United", make it easier to use money to influence outcomes. In 2014, Transparency International (www.transparency.org) ranked the U.S. as the 17th least corrupt out of 175 countries. Not all lobbying is corruption, but lobbying uses money and connections to influence outcomes.

An example illustrates lobbying and the collective action problem. Pollution reductions may impose costs on some groups, while benefitting society at large. We take the extreme case where all the costs of a policy fall on a group of firms, and all the benefits accrue to consumers. Both groups can lobby to influence the probability that this policy is implemented, and both groups face a collective action problem in financing their lobbying. The policy increases consumer welfare by 100 units and reduces firm welfare by 50 units, yielding a net benefit to society of 50 units. Absent lobbying, the policy is implemented with probability 0.5, so the expected benefit to society (the probability that the benefit occurs times the level of the benefit if it does occur) is 25. If only one group spends 10 units on lobbying, that group has its preferred outcome with certainty. If both groups spend 10 units on lobbying, their efforts cancel each other, leaving unchanged the probability that the policy is implemented, but wasting 20 units of welfare.

Table 1 shows the payoff matrix if each group is represented by a single agent who decides whether to lobby. The first element of each ordered pair shows consumers' expected payoff for a combination of actions, and the second element shows producers' payoff. If both groups lobby, consumers' expected payoff is $0.5\,(100) - 10 = 40$, and firms' expected payoff is $-0.5\,(50) - 10 = -35$, for a net social benefit of 5. This game illustrates the Prisoners' Dilemma. Each group is better off lobbying, regardless of what the other group does; lobbying is a "dominant strategy", and the only (Nash) equilibrium in this game is for both groups to lobby. However, both groups are better off (relative to the Nash equilibrium) if they forswear lobbying.

consumers\ firms	firms lobby	firms do not lobby
consumers lobby	$(40, -35)$	$(90, -50)$
consumers do not lobby	$(0, -10)$	$(50, -25)$

Table 91: Payoff matrix for the lobbying game. First element in an ordered pair shows consumers' expected payoff, second element shows firms' expected payoff.

This payoff matrix assumes that each group has solved its collective action problem, acting as a unified agent, i.e. the group has delegated authority to a single agent who decides, on the group's behalf, whether to lobby. However, the benefits of lobbying tend to be dispersed for consumers and concentrated for producers. Producer groups are legal for the purpose of representing the industry's interests to legislators; these groups can coordinate the individual firm's lobbying contributions. Firms are therefore more likely than consumers to solve their collective action problem.

For this example, society's expected payoff is 25 if neither group solves the collective action problem, so that neither lobbies. It is 5 if both groups solve their collective action problem, so that both lobby. Society's payoff is -10 if only firms solve the collective action problem. Here, society's payoff is highest if only consumers solve their collective action problem, and it is lowest if only producers solve their collection action problem; but consumers are less likely than producers to solve their collective action problem. A group's ability to solve its collective action problem need not benefit society.

Renewable Fuel Standard: an example of lobbying In 2005 the U.S. introduced a Renewable Fuel Standard (RFS), requiring annual minimum consumption levels of different biofuels; 2007 legislation increased these levels. The Environmental Protection Agency (EPA) implements the policy by estimating gasoline demand in the next year and dividing annual targets of the different biofuels by the estimated gasoline consumption, to obtain a ratio σ_i for biofuel i. Gasoline producers are required to use σ_i gallons of biofuel i for each gallon of gasoline they produce. These producers face a "blending constraint" that increases their cost of production, because the biofuels are more expensive than gasoline.

Proponents of the RFS justify it using an "infant industry" argument, claiming that biofuels will eventually be important both as low carbon alternatives to fossil fuels, and as alternatives to foreign sources of petroleum. Because the current state of technology and infrastructure would not enable this industry to survive under market conditions, government policy is needed to protect this "infant" until it grows into a mature industry. Infant industry arguments go back at least to the early 1800s, when they were used to justify trade restrictions. Many opponents of the RFS begin as skeptics of the infant industry argument, because of experiences where infants fail to mature. In addition, the applicability of the infant industry argument is questionable in this case, because the RFS has promoted the production of corn-based ethanol, for which the technology was already mature.[1]

[1] After 2015, ethanol produced using cellulosic material, including the inedible part of corn and special crops such as switchgrass, is scheduled to become more important in

The RFS emphasis on corn-based ethanol has three major disadvantages, in addition to having low potential to encourage new technology. First, it leads to a small, and by some estimates non-existent reduction in carbon emissions. Second, it diverts a major food crop from food to fuel, increasing food prices and worsening food insecurity in some parts of the world. Third, the policy has encouraged farmers to cultivate marginal land that would otherwise have been left fallow under a conservation program.

Recent evidence estimates that carbon reductions achieved using the RFS are about three times as costly as the reduction that could have been achieved under an efficient policy such as an emissions tax or cap and trade. An important consequence of the RFS was to provide large transfers from the general public (in the form of higher fuel prices) to corn growers, likely with little environmental or technological benefit. The RFS was estimated to increase U.S. fuel costs by $10 billion per year. Why did the U.S. government implement an inefficient policy instead of an efficient policy?

The (never-passed) Waxman-Markey bill (2009) would have imposed a cap on carbon emissions, and required that fuels eligible for the RFS produce greater carbon reduction than achieved at the time. Thus, Waxman-Markey would have reduced the transfers that corn producers receive under the RFS. Representatives tend to vote their constituents' interest. Representatives from districts that benefit under the RFS were more likely to oppose Waxman-Markey, and they also received greater campaign contributions from groups opposing the bill. The cap and trade policy under Waxman-Markey reduces emissions more efficiently than the RFS, but the gains from the latter are concentrated in a small number of districts, whereas the benefits of the former are widely dispersed. Lobbying opposed toWaxman-Markey received more financial support than lobbying favoring the bill.

9.4 Subsidies and the double dividend

Objectives and skills

- Understand why a tax and a subsidy have the same effect on a polluter's incentives.

the RFS. Cellulosic biofuels rely on an immature technology, where government support can potentially lead to large improvements. However, the RFS's support for corn-based ethanol is unlikely to promote the development of cellulosic biofuels.

- Understand the political and the economic disadvantages of a subsidy, compared to a tax.

- Know the outline of the double dividend hypothesis.

We consider the situation where a pollution externality creates a market failure, and some group is able to block a policy that would be socially beneficial. Here it makes sense to consider more costly, but politically acceptable, alternatives. An abatement subsidy is a plausible alternative to a pollution tax, but the cost of financing the abatement subsidy creates obstacles. This fiscal cost relates to the "double dividend hypothesis", an idea that appears to strengthen the argument for a pollution tax.

Subsidizing abatement

Instead of taxing firms for creating pollution, society can subsidize them for abatement (= reducing pollution). A firm has the same incentive to reduce pollution if it is taxed \$1 for each unit of pollution, or given a subsidy of \$1 for each unit that it abates. Facing the tax, a unit of pollution creates a direct cost to firms; facing the subsidy, a unit of pollution creates an opportunity cost to firms. These two costs have the same effect on the firm's incentives, so (in principle) the two policies achieve the same level of pollution.[2]

There are both political and economic obstacles to using the subsidy instead of the tax. The subsidy imposes a cost on taxpayers, requiring a transfer from general tax revenue to a specific group of firms. Even if the pollution reduction is worth this cost, it may be politically hard to convince voters to tax themselves to pay firms stop a socially harmful practice.

Raising revenue to finance subsidies creates a deadweight cost in addition to the direct distributional effect of taking income away from a group (Chapter 10). The distinction between a transfer and a deadweight cost is important. Taking \$1 from Mary to give to Jiangfeng is a transfer, not a cost to the economy. However, if the government has to take \$1.25 from

[2] Behavioral economics studies show that people's reservation price for selling an item (their "willingness to accept") frequently exceeds their reservation price for buying the same item (their "willingness to pay"). This asymmetry presents a type of "loss aversion", which is difficult to reconcile with perfect rationality. People who run firms may exhibit similar failures of rationality, causing them to respond differently to pollution taxes and abatement subsidies. However, the discipline of the marketplace encourages rationality.

Mary in order to give \$1 to Jiangfeng, there is a \$0.25 deadweight cost to the economy. The government makes transfers using a "leaky bucket".

The assumed fixed relation between pollution and output means that the only way to reduce one is to reduce the other; here, there is no deadweight cost in taxing the polluting sector. However, taxing other sectors to finance a subsidy for the polluting sector typically does create a deadweight cost. If the deadweight cost associated with general taxes is 25% of revenue raised by a tax, then financing a \$1 subsidy to this polluting industry creates a deadweight cost of \$0.25. If, instead, the polluting industry is taxed \$1, and that tax revenue transferred to general funds, then other taxes can be reduced by \$1.00, saving society the deadweight cost \$0.25. Replacing a \$1 pollution tax with a \$1 abatement subsidy increases social costs by \$0.50. Taxing pollution is likely more efficient than subsidizing abatement.

The double dividend

The TOSB alerts us to the possibility that a policy, such as the Pigouvian tax, that is optimal in the presence of a single distortion, may not be optimal when there are multiple considerations. The theory does not tell us whether considerations outside the polluting sector cause the optimal pollution tax to be above or below the Pigouvian level. The numerical example in the last paragraph illustrates the "double dividend hypothesis", an idea that implies that the optimal pollution tax exceeds the Pigouvian level. The tax lowers pollution (the first dividend) and by raising revenue it make it possible to reduce taxes in other sectors, lowering deadweight costs there (the second dividend). The Pigouvian tax addresses the goal of lowering pollution, but not the second dividend. The second dividend provides a reason to increase the pollution tax above the Pigouvian level.

An example outside economics illuminates the TOSB, and the double dividend in particular. A person who wants to get stronger may "target" an exercise regimen to this goal. Getting stronger corresponds to reducing pollution, and the targeted exercise regimen corresponds to the Pigouvian tax. If the person discovers that exercise also affects the difficulty of weight control, they might rethink their exercise regimen. The optimal change in this regimen depends on whether exercise makes it harder or easier to control weight. The general point is that once we take into account considerations other than our main objective (pollution reduction or getting stronger), we have to modify our policy/exercise plan.

Moving back to the economic context, suppose that the government requires a fixed amount of revenue, which it can raise using a combination of an income tax and a tax on the polluting sector. The double dividend hypothesis implies that the optimal pollution tax exceeds the Pigouvian level (equal to the marginal social cost of pollution). To test this hypothesis, we begin with a tax equal to the Pigouvian level; now consider the welfare effect of a perturbation that slightly increases this tax, making an offsetting change in the income tax to keep total tax revenue at the required level. Under reasonable (but not all) parameter values, analysis shows that this perturbation lowers welfare, thus rejecting the double dividend hypothesis.

The explanation for this counter-intuitive result begins with the fact that an income tax is more efficient than a commodity tax at raising government revenue.[3] Both taxes create deadweight losses, transferring money from private agents to the government using leaky buckets; but the commodity tax–bucket leaks more. The income tax drives a wedge between the price that workers receive, and the price that firms pay, for an hour of work. The tax reduces incentives to supply labor, causing firms to face a higher price of labor. That higher price discourages production across all (or most) sectors. The income tax therefore falls more broadly and evenly across the different sectors of the economy, creating a smaller effect on any individual sector, compared to a commodity tax. Absent the pollution externality, it is optimal to raise all of the necessary revenue using the income tax.[4]

Under the pollution externality, a perturbation that slightly *reduces* the pollution tax below the Pigouvian level creates only a small ("second order") welfare cost arising from increased pollution. Moving toward a more efficient tax structure (income instead of commodity taxes) creates a large ("first order") welfare gain. On balance, the perturbation increases welfare, "disproving" the double dividend hypothesis. The pollution tax creates a larger disincentive to supply labor, compared to the income tax. This analysis il-

[3] Under the assumption of a fixed relation between pollution and output of the dirty good, a commodity tax and a pollution tax are equivalent. Under more general assumptions, the two taxes differ, and the explanation offered here becomes more complicated.

[4] This claim assumes that the government is unable to use (non-distortionay) lump sum taxes. In the 1980s the British Prime Minister Margaret Thatcher attempted to reduce (distortionary) income and commodity taxes by introducing a poll tax (a particular lump sum tax). The effort was abandoned after sparking huge opposition; the poll tax was perceived as inequitable, falling most heavily on low income people. Economists emphasize efficiency over equity. Politicians have to be sensitive to equity.

lustrates “general equilibrium” relations: the pollution tax directly affects the polluting sector, and indirectly affects the labor market.

Why does the double dividend hypothesis fail here, when it seems so reasonable in the numerical example in the previous subsection? There are two parts to the answer. First, the analysis here assumes that the government can raise revenue using a (relatively) efficient income tax. Our numerical example assumed that the alternative means of raising revenue is a distortionary commodity tax. Second, the analysis here recognizes the general equilibrium effect of the pollution tax on the labor market; our numerical example uses a partial equilibrium framework, implicitly assuming that the pollution tax does not alter the distortion in the other sector, except to the extent that it decreases that distortion by decreasing the tax there. The first part of this answer is especially important in trying to reach a conclusion about the plausibility of the double dividend hypothesis. If the revenue from a pollution tax is used to decrease inefficient taxes, the double dividend hypothesis is plausible. If the revenue is used to offset relatively efficient taxes, the hypothesis is implausible.

9.5 Output and input subsidies

The welfare cost of distortions that reinforce each other can be much greater than the sum of the welfare costs of the distortions in isolation. Many agricultural markets involve both output and input subsidies. Both of those policies encourage excessive use of (some) inputs; the policies reinforce each other, leading to a combined welfare cost greater than their individual costs. Output and input subsidies can be explicit or implicit. Producers likely prefer implicit subsidies, because their lower visibility makes them easier to defend in the political arena. Explicit output or input subsidies pay producers a subsidy per unit of output produced or input purchased, creating transfers from taxpayers to producers. A trade restriction raises domestic price by limiting cheaper imports, providing an *implicit* output subsidy, creating a transfer from consumers and/or taxpayers to domestic producers. Implicit input subsidies arise, for example, if farmers’ water price is less than the full social marginal cost of water, equal to the cost of extracting and transporting the water plus its opportunity cost (the resource rent). This type of subsidy creates a transfer from taxpayers and future water users to farmers. U.S. sugar producers receive implicit output subsidies in the form of trade restrictions, and implicit input subsidies in the form of underpriced water (or unpriced pollution related to their water use).

These policies create transfers and distortions. The transfers have equity implications, but only the distortions matter from the standpoint of *efficiency*. Using the leaky bucket metaphor, the distortion corresponds to leaks in the bucket, the distortionary cost is the amount of water that leaks out, and the transfer is the amount of water that reaches the recipient.

An example illustrates the interaction between output and input subsidies. Under free trade a country can buy sugar at the world price, 1, and it can produce sugar, S, using labor, L, water, W, and a fixed input land, F (having no other uses); the production function is $S = F^{1-\alpha-\beta}L^{\alpha}W^{\beta}$. A trade restriction increases the domestic price of sugar to $1+s$; s is the trade-induced implicit subsidy to producers. The market for labor is efficient, with the price of labor equal to ω, its opportunity cost. The efficient price of water is p, but producers receive a subsidy, ϕ, so their cost for a unit of water is $p-\phi$. Water subsidies are often combinations of direct subsidies for the infrastructure required to extract and transport water, and an implicit subsidy caused by not charging users the efficient resource rent (Chapter 17).

Price-taking farmers hire labor and buy water to maximize profits,

$$\pi(s,\phi) \equiv \max_{L,W} \left[(1+s)\, F^{1-\alpha 1\beta} L^{\alpha} W^{\beta} - \omega L - (p-\phi)\, W\right]. \tag{9.1}$$

The difference between revenue and the payment to labor and water equals the rent (or profit), $\pi(s,\phi)$, earned by the owners of the fixed factor, F. The output and input prices (1, ω, p), are exogenous. For this experiment, we also fix the consumer price at $1+s$, thus fixing in consumer welfare.[5] In this setting, social welfare equals producer surplus (returns to the fixed factor, land), *excluding transfers* arising from the subsidies. The transfer increases producer welfare but creates an exactly offsetting welfare loss to agents who pay for it. Transfers are a wash from the standpoint of welfare.

The subsidy-induced misallocation of inputs creates a welfare loss. The (implicit) output subsidy, s, encourages farmers to produce too much sugar, causing them to buy too much labor and water, relative to the socially optimal level. The water subsidy, ϕ, causes farmers to buy too much water (and

[5] For example, Policy 1 allows free trade (so domestic producers face the world price) and taxes consumption at rate s, and Policy 2 imposes a unit tariff of s. These two policies have the same effect on consumers (causing them to face price $1+s$). Moving from Policy 1 to Policy 2 does not change consumer welfare, but raises producer prices from 1 to $1+s$. The second policy provides a production subsidy and the first does not.

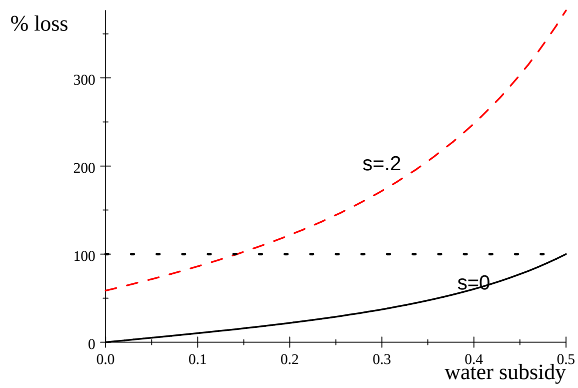

Figure 9.2: The solid graph shows the percent loss in welfare, as a function of the water subsidy, when the output subsidy is 0 ($s = 0$). The dashed curve shows the percent welfare loss when the output subsidy is 20% ($s = 0.2$).

hire too little labor, *conditional on their output level*). The two subsidies reinforce each other: the distortion caused by the water subsidy is worse, the larger is the output subsidy. Figure 9.2 shows the percent loss in welfare due to the water subsidy, for two values of the output subsidy. The figure illustrates:

- With $s = 0$, the welfare cost of the water rises slowly with the water subsidy, ϕ: small subsidies create small losses.
- The welfare cost of ϕ for $s = 0.2$ (and for any $s > 0$) rises rapidly with the water subsidy. With $s > 0$, even a small water subsidy creates a large *additional* welfare loss.
- A sufficiently high water subsidy causes the welfare loss to exceed 100%. There, sugar production has negative value added.

The return to land (the farmer's profit) is always positive in this setting. Some of that profit derives from the transfers discussed above. When the welfare cost exceeds 100%, the social value of labor and water exceeds the social value of sugar production; all of the farmer's profits derives from the transfers. In this case, sugar production lowers social welfare; shutting down the industry would raise social welfare, even though it means idling cropland.

Calculating subsidies' welfare cost Denote $W(s,\phi)$, $S(s,\phi)$, and $\pi(s,\phi)$ as, respectively, the water purchases, sugar production, and farmer profit, all functions of the subsidies. We find these by (numerically) solving the first order condition to the problem in equation 9.1. Zero subsidies lead to efficient production and the maximum level of social welfare, $\pi(0,0)$.

Under non-zero subsidies, producer surplus equals $\pi(s,\phi)$, and the transfer equals $T(s,\phi) \equiv sS(s,\phi) + \phi W(s,\phi)$. The social value of each unit of sugar production equals the world price of sugar, 1. (An extra unit of domestic production saves society the cost of importing one unit, which costs \$1 of foreign exchange.) The farmer who receives the price $1+s$ and produces S units of sugar receives a transfer sS. This amount equals farmer revenue in excess of the social value of production. The social cost of each unit of water equals p. The farmer who buys W units of water at price $p-\phi$ receives a transfer of ϕW. Both parts of the transfer increase farmer profit, but they cause an exactly offsetting loss in welfare to other agents, e.g. taxpayers. Therefore, in calculating social welfare under subsidies, we have to "net out" the transfers created by the subsidies. Social welfare under the subsidies equals $\pi(s,\phi) - T(s,\phi)$. The social cost of the subsidies, as a percent of the optimal level of welfare, equals

$$\frac{\pi(0,0) - [\pi(s,\phi) - T(s,\phi)]}{\pi(0,0)} 100.$$

Figure 9.2 graphs this ratio as a function of ϕ for $s=0$ and $s=0.2$, using the parameter values $\alpha = 0.6$, $\beta = 0.2$, $\omega = 1 = p$.

9.6 Policy complements

- Examine two market failures through the lens of the TOSB.
- Determine the socially optimal pollution tax when marginal costs increase with the level of pollution.
- Determine whether two policies are complements or substitutes.

Positive research spillovers, where research conducted by one firm helps other firms, create a rationale for green industrial policy. Green subsidies are also often supported as a second best alternative to politically infeasible policies. The American Enterprise Institute and the Brookings Institute, a conservative and a liberal think tank, respectively, both endorsed R&D subsidies for green technology as an alternative to carbon taxes. These green policies subsidize firms for doing something socially useful (developing new technologies) instead of for refraining from doing something socially harmful (polluting), and therefore encounter less political resistance than an abatement subsidy.

Policy substitutes or complements

Is Green industrial policy really an alternative to, i.e. a *substitute* for emissions taxes (as they appear to be), or are they *complements* to emissions taxes? The terms complements and substitutes are familiar from demand analysis. Two policies are said to be substitutes if the implementation of one makes the other less valuable to society; they are complements if the implementation of one makes the other more valuable.

Chapter 8 shows that green industrial policies might aggravate the pollution problem. Here we develop the closely related idea, that instead of being a substitute for a carbon tax, green industrial policy might make the carbon tax more, not less, vital to society: the policies may be complements.

In a competitive resource (e.g. fossil fuel) sector,

$$p_t - \frac{\partial c\left(x_t, y_t\right)}{\partial y} = R_t, \text{ or } p_t = \frac{\partial c\left(x_t, y_t\right)}{\partial y} + R_t, \tag{9.2}$$

where marginal extraction cost equals $\frac{\partial c(x_t,y_t)}{\partial y}$; R_t equals the firms' period-t rent, the opportunity cost of extraction in period t. Thus, $\frac{\partial c(x_t,y_t)}{\partial y} + R_t$ is the "combined" marginal extraction cost, including both the standard marginal cost and the opportunity cost. Chapter 5.2 shows that R_t equals the firms' present discounted value of *future* rents, plus any cost reduction due to a higher stock. Chapter 8 explains why green industrial policy might reduce future rents, thus reducing the firms' period-t rent, R_t.

Increasing marginal costs of pollution

As with the example in Chapter 9.2, we assume that one unit of production (here, extraction) creates one unit of pollution: reducing pollution requires a corresponding reduction in output. In the previous example, we took the marginal damage resulting from pollution to be a constant, but here we assume that marginal damage increases with the level of extraction. Extraction (= emissions, by choice of units) of y creates damages equal to y^2, so marginal damages equal $2y$. Where marginal damage are constant, the Pigouvian tax equals that constant. Here, however, where marginal damages depend on the level of extraction, the Pigouvian tax also depends on the level of extraction. We find the Pigouvian tax by:

1. Identifying the socially optimal level of extraction, equating price to social marginal costs (= marginal extraction costs, plus the opportunity cost, R_t, plus marginal damage, $2y_t$.)

2. Once we have found the socially optimal level of output, denoted y_t^*, we set the Pigouvian tax equal to the marginal external cost, $2y_t^*$ (for this example).

The effect of rent on the Pigouvian tax

Figure 9.3 illustrates the case where the firm has constant marginal extraction costs, $\frac{\partial c(x_t,y_t)}{\partial y} = C$. Suppose that at a point in time the private combined marginal cost is $C + R_t = 10$, shown by the solid flat line in the figure. For our example, the social marginal cost equals the private marginal cost plus the (external) environmental marginal damage, $10+2y$. If a green industrial policy lowers resource rent by five units, the *private* combined marginal cost, and also the social marginal cost, falls by 5 units.

The upwardly sloping solid curve in Figure 9.3 is the graph of $2y+C+R_t$, the *full* social marginal cost, equal to marginal damage plus the private cost. A five-unit reduction in rent causes this combined social marginal cost to shift from the positively sloped solid line to the dashed line. This reduction in social cost increases the socially optimal level of production from the intersection shown at point A, to the intersection shown at point B.

The reduction in rent decreases the combined social cost of extraction, thereby increasing the socially optimal level of extraction $\Rightarrow$ increasing marginal damages $\Rightarrow$ increasing the optimal tax. This result is general: where marginal damages increase with the level of extraction (or production), a decrease in firms' marginal cost leads to an increase in the optimal tax. When firms' marginal cost falls, they tend to produce more, and the higher production leads to higher marginal damages and a higher optimal tax.

The Pigouvian tax induces the firm to face the social cost of production, giving firms the correct incentive to produce at the socially optimal level. If

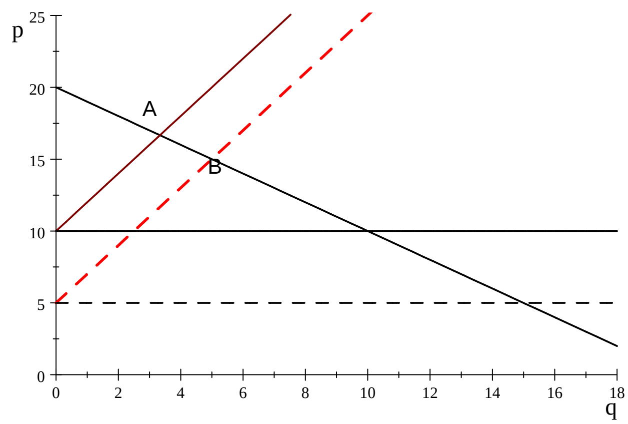

Figure 9.3: External marginal costs = $2y$ (or $2q$). Solid positively sloped curve shows social marginal costs when private marginal cost $C + R_t = 10$. Socially optimal production occurs at A. If private marginal costs fall to $C + R_t' = 5$, socially optimal production occurs at B. When private costs fall, the optimal tax increases.

the firm faces a constant tax ν and has private costs (inclusive of opportunity cost, its rent) equal to 10, its private cost equals social costs if and only if the tax equals the vertical distance from point A to 10. The Pigouvian tax thus equals this distance, which we denote ν^A. If the firm's rent falls so that its rent-inclusive private cost now equals 5, the Pigouvian tax equals the vertical distance from point B to the flat dashed line; we denote this tax as ν^B. It is apparent from the figure that $\nu^B > \nu^A$. The decrease in rent increases the optimal tax. In this case, an emissions tax and a green industrial policy that reduces rent are *complements*, not *substitutes*.

The algebra Figure 9.3 uses the inverse demand (equal to the marginal benefit of consumption) $p = 20 - y$. The socially optimal level of production, y^*, equates the marginal benefit of consumption (the market price) to the full social cost of production, equal to the private cost, $C + R$ $(= 10)$, plus the externality cost, $2y$. This equality requires $20 - y = C + R + 2y$, or $y^* = \frac{20-R-C}{3}$. A competitive firm facing the tax ν produces where the price equals its private cost plus the tax, implying $20 - y = C + R + \nu$, or $y^{\text{compet}} = 20 - R - \nu - C$. The Pigouvian tax causes the competitive firm to produce at the socially optimal level, requiring $y^{\text{compet}} = y^*$, or $\frac{20-R-C}{3} =$

$20 - R - \nu - C$, implying $\nu^{\text{Pigouvian}} = \frac{40}{3} - \frac{2}{3}R - \frac{2}{3}C$. A reduction in rent (leading to an outward shift in the firm's supply function), increases the Pigouvian tax. In this example, green industrial policy lowers the resource rent, increasing the Pigouvian tax: the green industrial policy and the carbon tax are policy complements.

Relation between this example and the TOSB

This tax example illustrates the possibility that policies that appear to be either unrelated or substitutes might be, on closer examination, complements. The TOSB reminds us that connections that are not apparent may nevertheless be important. Green industrial policy might make carbon taxes more, not less important. The TOSB warns that in the presence of two or more distortions, or market failures, correcting only one of those distortions might exacerbate the other distortion to such an extent that welfare falls.

9.7 Summary

This chapter introduces the notion of a second-best policy or outcome: one that is not optimal, or "first best". The Principle of Targeting recognizes the importance of carefully matching policies and objectives. A Pigouvian tax causes competitive firms to internalize an externality. For example, a Pigouvian tax causes firms to take into account the social cost of pollution when making production decisions. Two policies are said to be substitutes if the implementation of one policy makes the other policy less important, or decreases the optimal level of the other policy; they are complements if the implementation of one policy makes the other more important, or increases the optimal level of the other policy.

The theory of the second best (TOSB) and the Principle of Targeting (POT) are deceptively simple ideas, with important economic implications. The TOSB reminds us that in economies, "the hip bone is connected to the shoulder bone", although perhaps not directly. Because markets connect apparently unconnected outcomes, a policy that reduces one market failure may, in the presence of a second market failure, actually lower welfare. We illustrated this result using an example of a monopoly that produces pollution; moving from the zero emissions tax to the Pigouvian tax might decrease social welfare. A policy that is optimal under perfect competition might be harmful under monopoly. The POT reminds us to beware of collateral damage in setting policy, and to attempt to select policies that do not inflict such damage.

We defined the collective action problem, and used the Renewable Fuel Standard to illustrate political reasons for the adoption of inefficient policies. The deadweight costs associated with raising government revenue tend to make pollution taxes more efficient than abatement subsidies. The double dividend hypothesis claims that the pollution tax permits a reduction in other taxes, causing the optimal pollution tax to exceed the Pigouvian level. This hypothesis is likely true if the taxes being replaced are relatively inefficient, and likely false if those taxes are efficient.

Taxes and subsidies "distort" equilibrium allocations. Two policies might either reinforce or offset each other; in the former case, the welfare cost of the policies taken together exceeds the sum of their costs in isolation. An example showed that multiple distortions might cause a sector to have negative value added, making the social value of the sector's output less than the social value of the inputs used in the sector.

The TOSB makes us careful about the application of economic intuition. Common sense might suggest that a politically palatable policy (e.g. a green subsidy) is a substitute for, or alternative to, a politically difficult policy (e.g. a carbon tax). More careful analysis may reveal that such policies are complements: the green subsidy makes the carbon tax more, not less, important. Much of our economic intuition is developed from studying simple models with a single distortion. The real world is more complicated. Economic analysis is closer to chess than to checkers.

9.8 Terms, study questions and exercises

Terms and concepts

Theory of the second best, Principle of Targeting, Pigouvian tax, a tax "supports" or "induces" an outcome, internalize an externality, collective action problem, payoff matrix, Nash equilibrium, Prisoners Dilemma, Renewable Fuels Standard, deadweight loss, abatement, double-dividend, increasing marginal pollution costs, policy complements and substitutes.

Study questions

1. Use a graphical example (and a static model) to show that the Pigouvian tax that corrects a production-related externality (e.g. pollution) in a competitive setting might lower social welfare if applied to a monopoly.

2. (a) Consider a static model. Suppose that inverse demand is $10 - q$, firms are competitive with constant average = marginal costs C, and pollution-related damages (arising from output) are $q + \frac{1}{2}\beta q^2$, with $\beta > 0$ What is marginal damage? (b) What is the socially optimal level of production and consumer price, and what is the Pigouvian tax? (c) How does the Pigouvian tax change with β? (d) Provide the economic explanation of this relation.

3. (a) Continue with the model in question #2. Suppose that there is a policy that reduces C, e.g. by making production more efficient. How does this reduction in C alter the Pigouvian tax that you identified in question #2b? (b) Are the two policies (the pollution tax and the policy that reduces C) complements are substitutes? (c) How would the answer to part (b) have changed if $\beta = 0$? Explain.

4. You are in a conversation with someone who correctly states that, in a particular market, international trade increases production in poorer countries with weaker environmental standards, thereby increasing a global pollutant (i.e. a pollutant that causes worldwide damage, not just damage in the location where production occurs). The person claims that a trade ban is a good remedy for this problem. Regardless of your actual views, use concepts from this chapter to argue against this person's proposal.

5. (a) Explain why, in principle, a tax on pollution and a subsidy to abatement have the same consequences for society. (b) Summarize the political and the economic reasons why in practice, an abatement subsidy and a pollution tax are likely to have different consequences for society.

6. Describe (in a few sentences) the "double dividend hypothesis" and the rationale for the hypothesis.

Exercises

1. Suppose that consumers of the product bear the cost of pollution. The model then contains three types of agents: the firm, consumers, and taxpayers. Taxpayers benefit from tax revenue and they do not like having to pay the cost of subsidies. Inverse demand is $p = 20 - 3q$, private average = marginal cost is 2, and environmental damage per unit of output is 6. A monopoly chooses the level of sales. Using a figure like Figure 9.1 identify graphically (by shading in appropriate areas) the change in welfare of the three types of agent when a regulator imposes a unit tax of $\nu = 6$.

2. For this example, identify (graphically) the socially optimal tax/subsidy under the monopoly. That is, identify the tax/subsidy that induces the monopoly to produce at the socially optimal level.

3. Change the example in Exercise 1 to $p = 20 - 0.4q$. (Replace the slope 3 by 0.4) Other parameter values are unchanged. (a) Does this change make demand more or less elastic? (b) Find the optimal pollution tax for the competitive firm. (c) Find the optimal pollution tax under the monopoly. (d) Provide an economic explanation for the relation between the optimal tax and the slope of the inverse demand function, under both a competitive firm and a monopoly.

4. Suppose that (private) constant marginal costs is c, each unit of pollution creates d dollars of social cost (external to the firm), and the demand function is demand $= Q(p)$. (a) What is the optimal pollution tax for the competitive industry? (Here you do not have a specific functional form for demand, so your answer involves the function $Q(p)$, not a number.) (b) Use the formula for marginal revenue (a function of price and elasticity of demand) to find the equation for the optimal tax (or subsidy) under a monopoly. (c) Now suppose that $Q = p^{-\eta}$ with $\eta > 1$. Use your formula from part (b) to find the optimal tax/subsidy under the monopoly. (d) Provide an economic explanation for the relation between the optimal tax/subsidy under the monopoly and η.

5. Suppose that demand is $p = 20 - q$, private constant marginal production cost equals 10, and marginal environmental damages equal $2 + \alpha q$, where the parameter $\alpha \geq 0$. Firms are competitive. (a) Discuss the

economic interpretation of the parameter α. In particular, explain the difference in the model with $\alpha = 0$ and with $\alpha > 0$. (b) Find the socially optimal level of production when private costs equal 10. Find the optimal (Pigouvian) tax in this case. (Both of these are functions of α.) (c) Now suppose that private costs fall to 5; find the socially optimal production level and the Pigouvian tax with these lower private costs. (d) Write the difference in the Pigouvian tax under the high and the low private cost, as a function of α. (e) Describe and explain the effect of α on the change in the Pigouvian tax arising from the change in private costs.

6. *Description of setting.* Inverse demand in the polluting sector is $p = 20 - 3q$, private average = marginal cost is 2, and environmental damage per unit of output is 6. In Scenario A the government is able to raise revenue without creating any distortion (e.g. by means of a lump-sum tax). In Scenario B, the economy-wide average of deadweight loss from taxes is 10% of the tax revenue. This assumption means that an extra \$1 in government revenue raised using a non-distortionary tax (one with zero deadweight loss) is worth \$1.10, because that revenue makes it possible to maintain the same level of public expenditure while reducing tax from the distortionary source. Here, the social value of tax revenue \$$TR$ raised from a non-distortionary source is \$$1.1TR$. In both scenarios, raising revenue by taxing the polluting sector creates no deadweight cost. *The question:* What is the optimal Pigouvian tax for a competitive industry in Scenario A and what is the optimal tax in Scenario B? Explain their relative magnitudes. (If you are unable to answer this question using mathematics, use the discussion in Chapter 9.4 to provide a qualitative answer.)

7. Does the positive social value of government revenue (10% here) increase or decrease the optimal tax in the polluting sector? (Explain the qualitative effect, on the optimal tax under competition, of the positive social value of tax revenue (10% in our example). (You should use economic logic – not math – to figure out whether the policy under each market structure gets larger or smaller (in absolute value) when there is a cost to public funds. There is nothing tricky about this question; you just have to use "common (economic) sense".

Sources

Lipsey and Lancaster (1956) is the classic article on TOSB.

Okum (1975) is credited with the "leaky bucket" metaphor.

Fowlie (2009) and Holland (2012) provide recent applications of the TOSB related to environmental regulation.

Tritch (2015) discusses the statistics on U.S. lobbying provided by the Sunlight Foundation.

Auerbach and Hines (2002) survey the literature on taxation and efficiency.

Prakash and Potoski (2007) provide examples of collective action in environmental contexts.

Bovenberg and van der Ploeg (1994) and Goulder (1995) and Bovenberg (1999) discuss the double-dividend hypothesis.

Winter (2014) shows that carbon taxes and green industrial policy are likely to be policy complements.

Leonhardt 2010 describes the political popularity of green industrial policy.

Holland et al. (2015) study the political economy connections between the RFS and the Waxman - Markey bill.

The book written by a committee convened by the National Research Council (2011) discusses U.S. Biofuels policy.

Bryce (2015) provides the estimated cost to motorists ($10 billion/year) of the RFS.

The USDA website http://www.ers.usda.gov/topics/crops/sugar-sweeteners describes US sugar policy.

Chapter 10

Taxes: an introduction

Objectives

- Understand taxes' effect on market outcomes, and the principles of taxation.

Information and skills

- Understand the definition of tax incidence and be able to explain why a tax on consumers or on producers are "equivalent" in a closed economy.
- Understand the relation between tax incidence and supply and demand elasticities.
- Identify tax-induced changes in consumer and producer surplus, and identify the deadweight cost of a tax.
- Understand the relation between "cap and trade" and a pollution tax; apply intuition about taxes to study cap and trade.

Understanding the effect of taxes in the familiar static model of a competitive firm is worthwhile for its own sake, and also necessary for understanding the effect of taxes applied to natural resources, studied in Chapter 11. We emphasize competitive *closed* markets: one without international trade in the taxed commodity. This assumption means that domestic supply equals domestic demand.[1]

[1] Appendix F contains technical material including: algebraic verification of tax equivalence in the closed economy; an example showing that tax equivalence does not hold in an open economy; details on the approximation of tax incidence, deadweight loss, and tax revenue, and details related to the material on cap and trade.

Taxes raise revenue, making it possible to reduce other taxes while funding the same level of government expenditure. Taxes alter consumer and producer behavior, changing the equilibrium price and quantity, and producer and consumer surplus. There are three types of agents in our model, consumers, producers, and taxpayers; many people belong to two or all three of these groups. Consumer surplus, producer surplus, and tax revenue measure these agents' surplus. Social welfare equals the sum of the three measures.

10.1 Tax incidence and equivalence

Objectives and skills

- Introduce and define a unit tax and an ad valorem tax.
- Understand the meaning of tax incidence and tax equivalence.
- Explain tax equivalence in a closed economy.

If the government imposes a "unit tax" of $\nu = \$6$, and producers receive $\$p$ per unit sold, then consumers must pay $\$p + 6$ per unit. The difference between the consumer and producer prices equals the unit tax. An ad valorem tax, denoted τ, is measured as a rate. If the tax rate is τ and producers receive $\$p$ per unit sold, then consumers pay $\$(1+\tau)\,p$ per unit. There is a simple relation between the unit and the ad valorem tax. If producers receive the price p and one group of consumers pays a unit tax ν and another group of consumers pays an ad valorem tax τ, the two groups pay the same price if $p + \nu = (1+\tau)\,p$. Thus, two taxes yield the same consumer price if and only if $\nu = \tau p$. We can work with whichever type of tax we want, and easily translate one type of tax into another.

We assume that people are "rational", in the sense that their willingness to buy a commodity depends on the price they pay, not the precise manner in which the price is calculated. For example, a rational consumer is just as likely to buy a commodity priced at $1.10 "out the door" as a commodity marked at $1.00 that requires payment of a 10% sales tax at the cash register. In both cases, the final price equals $1.10. Behavioral economics shows that people sometimes react differently in these two settings.

Tax incidence and tax equivalence It might seem that it matters whether a tax is levied on consumers or producers (on which group has the "statutory obligation" to pay the tax). However, in a closed economy, the equilibrium price and quantity, and thus the consumer and producer surplus and the tax revenue, are the same regardless of whether the tax is levied on consumers or producers: consumer and producer taxes are "equivalent". Suppose that in the absence of a tax, the equilibrium price is \$12 and the equilibrium supply = demand is 100 units. Now consider a tax of \$2 per unit imposed on consumers. Does the imposition of this tax mean that the price consumers pay rises to \$12+\$2=\$14? In general, the answer is "no". The tax does increase the price that consumers pay, but (in general) this higher price decreases the amount that they demand. In order for producers to want to decrease the amount that they supply, the price that producers receive must fall. The increase in consumer price, as a percent (or fraction) of the tax is called the consumer incidence of the tax, and the decrease in producer price, as a percent (or fraction) of the tax is the producer incidence.

If the \$2 tax causes the tax-inclusive price that consumers face to rise from \$12 to an equilibrium of \$13.50, then the price that producers receive equals $13.5 - 2 = 11.5$, because the difference between consumer and producer price always equals the unit tax. Consumers "effectively" pay the share

$$\frac{13.5 - 12}{\text{tax}} = \frac{1.5}{2} = 0.75,$$

or 75% of the tax, and producers "effectively" pay the remaining 25% of the tax. The tax incidence on consumers is 75% and the tax incidence on producers is 25%. The tax incidence depends on the elasticities of supply and demand, but *not* on which agent has the statutory obligation to pay the tax. This equivalence between the producer and consumer taxes arises because, in a closed economy (no international trade) domestic production (supply) equals domestic consumption (demand).

Taxing polluters or "pollutees" The equivalence of producer and consumer taxes (in a closed economy) implies that it may not matter whether an externality is corrected using a tax on production or on consumption. That equivalence undercuts the advice that polluters (instead of those who suffer from pollution, the "pollutees") pay the cost of pollution: the "Polluter Pays Principle". If consumers are a proxy for the agent that suffers from the pollution, the principle implies that it matter whether the tax is levied on consumption or production. However, the equivalence between these two taxes implies that it is immaterial which tax is used. In this sense, the Polluter Pays Principle sets up a meaningless distinction. The larger point is that an environmental policy that raises production costs, affects both consumers and producers.

Consider the case where each unit of production creates \$2 worth of environmental damage, external to the firm. We also assume that the environmental damage is an inevitable consequence of production. Production and pollution are equivalent: society cannot have one without the other. (See the Caveat at the end of Chapter 9.2.) The optimal policy causes firms to internalize this environmental cost, just as they internalize costs associated with hiring capital and labor. A \$2 per unit producer tax achieves this goal, but in view of the equivalence of a producer and consumer tax, so does a \$2 consumer tax. The incidence of the two taxes is the same and they have the same effects on: the level of environmental damage, tax revenue, and consumer and producer surplus. It does not matter whether polluters (producers) or "pollutees" (consumers, as proxies for society) face the statutory obligation to pay the tax.

It also does not matter which agent is responsible for the environmental damage. Driving, a major source of environmental damage arises from consumption of the good (cars) rather than production. Suppose that production causes no pollution, but that each unit of consumption causes \$2 worth of environmental damage and, as in the previous example, there are no opportunities for abatement apart from reducing consumption. The optimal policy charges consumers a consumption tax equal to the marginal cost of pollution. In view of the equivalence between producer and consumer taxes for *non-traded* goods, we obtain the same outcome by imposing the statutory tax obligation on producers.

10.2 Tax incidence and equivalence (formal)

Objectives and skills

- Use graphs to show how a tax causes a shift in demand or supply, thereby identifying the effect of a tax on price and output.
- Use graphs and algebra to show the equivalence of consumer and producer taxes.

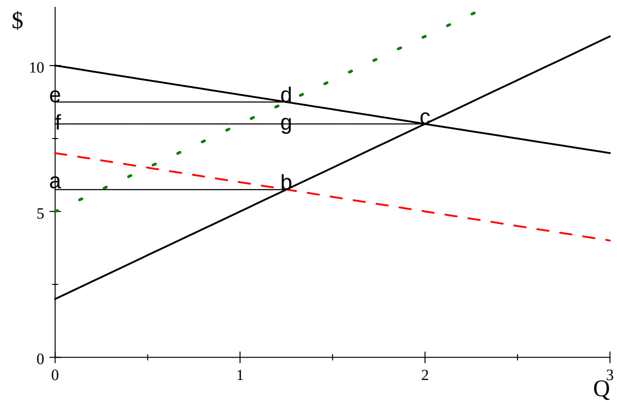

Figure 10.1: Solid curves show supply and demand absent the tax. Dotted curve shows supply curve as a function of consumer price (the producer has statutory obligation to pay tax). Dashed curve shows demand curve as a function of producer price (the consumer has statutory obligation to pay tax)

For "rational" agents, consumer and producer taxes are equivalent in a closed economy. Figure 10.1 shows supply and demand curves (heavy lines) without taxes; the equilibrium price and quantity is at point c, where consumers and producers face the same price. Once we introduce a tax, the consumer and producer prices are different, so we can no longer use the same axis to measure both prices. We have to be clear about what the vertical axis now measures. Suppose that we introduce a consumer unit tax of ν. We continue to let the vertical axis be the price that producers receive and we continue to denote the producer price by p. Therefore, the tax does not alter the location of the supply curve. The tax causes the consumer price to be $p + \nu$. The original demand function, the solid downward sloping line, shows the relation between quantity demanded and the *price that consumers pay*. However, under the tax we decided to use the vertical axis to represent the price that producers receive. Since the price that consumers pay and the price that producers receive are not the same when a tax is imposed, we cannot use the original demand function to read off the quantity demanded for an arbitrary producer price. The difficulty is that supply is a function of p and demand is a function of $p + \nu$, and we cannot let one axis represent both of these values. This difficulty is easily resolved.

The demand function for consumer tax ν equals the original demand function, shifted down by the magnitude ν, leading to the dashed demand function in Figure 10.1 The vertical distance between the original demand function and the demand function under the tax, is ν. This "new" demand function shows demand as a function of the producer price rather than the consumer price.

The intersection of the original supply function and the new demand function occurs at point b, showing the equilibrium quantity and producer price under the tax. The equilibrium consumer price (at point d), equals the producer price plus ν. Denote the distance between any two points x and y as $\|xy\|$. The tax *increases* the consumer price by $\|gd\|$ and *decreases* the producer price by $\|bg\|$. The sum of these two changes is $\|bd\| = \nu$. The consumer and producer taxes incidences are $\frac{\|gd\|}{\tau}100\%$ and $\frac{\|bg\|}{\tau}100$, which sum to 100%.

The paragraphs above assume that consumers bear the statutory obligation of paying the tax, so the tax shifts the demand function. If instead, producers bear the statutory obligation of paying the tax, then we let the vertical axis represent the price consumers pay. In this situation, the tax does not change the demand function, but it causes the supply function to shift up by ν units, as shown by the dotted supply function. It is apparent from Figure 10.1 the equilibrium quantity and the tax-inclusive consumer and producer prices are the same, regardless of which agent has the statutory obligation to pay the tax: consumer and producer taxes are equivalent

Algebraic example Suppose that inverse demand is $p = 10 - Q$ and marginal cost (= inverse supply) is $MC = S = 2 + 3Q$. In the absence of a tax, setting supply equal to demand implies the equilibrium price $p^* = 8$ and the equilibrium quantity $Q^* = 2$. If consumers have the statutory obligation to pay a unit tax $\nu = 3$, inverse demand, written as a function of the producer price, p, shifts to $10 - Q - 3$ (because consumers have to pay $p + 3$); the supply function is unchanged, so equilibrium occurs where $10-Q-3 = 2+3Q$, implying the equilibrium quantity $Q = 1.25$, the producer price $2 + 3\,(1.25) = 5.75$ and the consumer tax-inclusive price $10 - 1.25 = 8.75$. The consumer tax incidence is $\frac{8.75-8}{3}100 = 25\%$ and the producer tax incidence is $\frac{8-5.75}{3}100 = 75\%$.

If producers have the statutory obligation to pay the tax, inverse demand (as a function of the consumer price) remains at $p = 10 - Q$, but the supply curve shifts to $S = 2 + 3Q + 3$ (because producers deduct the tax from the payment they receive from consumers). Setting the demand equal to supply gives $10 - Q = 2 + 3Q + 3$, or $Q = 1.25$, as above. The consumer and producer prices, and therefore the tax incidences, are also the same.

10.3 Tax incidence and deadweight cost

Objectives and skills

- Determine how supply and demand elasticities affect tax incidence.
- Identify graphically the deadweight cost of a tax and show its dependence on supply and demand elasticities.
- Understand the difference between short and long run elasticities, and the resulting "time-consistency" problem.

Calculating the exact tax incidence requires that we find the equilibrium price in the absence of the tax, and the equilibrium consumer (or producer price) under the tax, and compare the two. (Using the fact that the tax incidences sum to 100%, we easily find one tax incidence by knowing the other.) We can use supply and demand elasticities to approximate the tax incidence for small taxes. The elasticities of supply and demand, evaluated at the *equilibrium price in the absence of a tax* are

$$\begin{aligned} \text{elasticity of supply } \theta &= \tfrac{dS(p)}{dp}\tfrac{p}{S} \\ \text{elasticity of demand } \eta &= -\tfrac{dD(p)}{dp^c}\tfrac{p}{D}. \end{aligned} \tag{10.1}$$

The change in equilibrium price due to a change in the tax, starting from a zero tax is

$$\frac{dp}{d\nu} = -\frac{\eta}{\theta + \eta}. \tag{10.2}$$

Equation 10.2 is a "comparative static expression" (Chapter 2.2). Equation 10.2 and the elasticity definitions produce approximations of producer and consumer tax incidence:

$$\begin{aligned} \text{producers' approx. tax incidence: } &\tfrac{\eta}{\theta+\eta}100 \\ \text{consumers' approx. tax incidence: } &\tfrac{\theta}{\theta+\eta}100. \end{aligned} \tag{10.3}$$

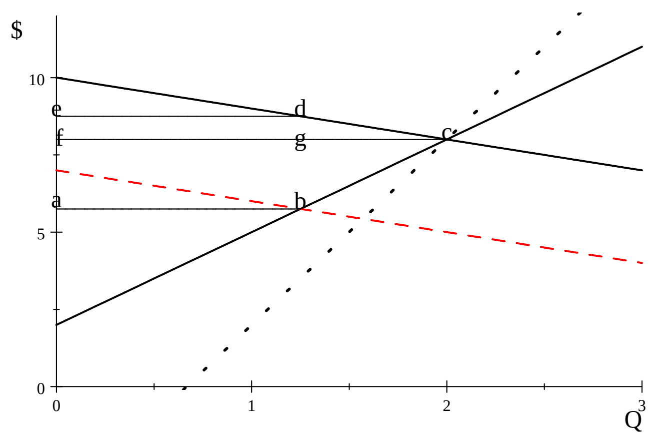

Figure 10.2: A less elastic supply (the dotted instead of the solid supply curve) increases the producer incidence of the tax

A lower elasticity of supply corresponds to a steeper supply function and a larger producer tax incidence. Figure 10.2 reproduces Figure 10.1, showing the original demand and supply functions, and the effect of a consumer tax. The figure includes a steeper (less elastic) supply function, the dotted line. Readers should identify the equilibrium quantity and the consumer and producer prices under the tax, to show that the less elastic supply curve increases the producer incidence of the tax. Similarly, smaller values of η mean that demand is less elastic, implying a steeper inverse demand function and a higher consumer tax incidence. By rotating the demand function around point c, readers can visualize the effect of making demand less elastic.

The trapezoid $fcde$ in Figure 10.1 measures the loss in consumer surplus due to the tax, and the trapezoid $abcf$ is the loss in producer surplus. Tax revenue equals the rectangle $abde$. Social welfare is the sum of producer and consumer surplus and tax revenues. The reduction in social welfare, due to moving from a zero tax to a positive tax, equals the reduction in consumer and producer surplus, minus the increase in the tax revenue. In Figure 10.1, this net loss equals the triangle bcd, society's deadweight loss (DWL) of the tax. The distortionary cost of the tax is small relative to the size of the transfer from consumers and producers to taxpayers. ("Triangles are small relative to rectangles.") In the case of linear supply and demand functions, the DWL is literally a triangle (known as the "Harberger triangle"). For

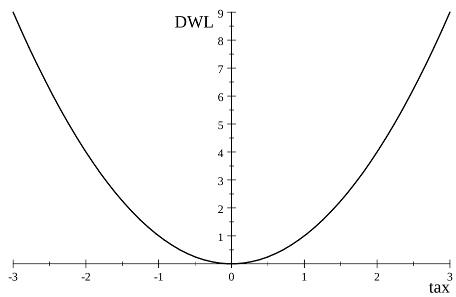

Figure 10.3: The graph of the approximation of deadweight loss.

general supply and demand functions, the approximate DWL is:

$$DWL \approx \left(\frac{1}{2} \frac{\theta\eta}{\theta + \eta} \frac{q}{p} \right) \nu^2. \tag{10.4}$$

The formula shows

> Result (i): The DWL is approximately proportional to the *square* of the tax. Result (ii): The deadweight cost is lower, the smaller is the elasticity of supply or demand.

It is not surprising that the DWL is zero for a zero tax and increases with the magnitude of the tax; the more important point is that it increases faster than the tax (Result i). Figure 10.3 illustrates this relation, showing that the DWL is a convex function of the tax. This fact implies that it is efficient to use a broad tax basis. For example, we may be able to raise the same amount of revenue by using a tax $\frac{\nu}{2}$ in each of two markets, instead of a tax of ν in a single market. Denote the term in parenthesis in equation 10.4 as X. If X is the same for both markets in our example, then the deadweight cost of using the tax ν in one market is approximately $X\nu^2$, whereas the deadweight cost of using $\frac{\nu}{2}$ in the two markets is approximately $2 \times X \left(\frac{\nu}{2}\right)^2 = \frac{X}{2}\nu^2$. For this example, doubling the tax base reduces the deadweight cost by 50%. Result (ii) implies that, other things equal, a tax applied to a commodity with low elasticity of supply or demand reduces the efficiency cost of the tax. For emphasis, we repeat the two "rules" of tax policy:

> (1) It is better to have a broad tax base (i.e. tax many instead of few goods). (2) It is better to tax goods that have lower elasticity of supply or demand.

A third "rule" is obvious: It is better to tax "bads", such as pollution, rather than "goods", such as labor or investment. Many taxes ignore some or all of these rules.

Product markets and general equilibrium analysis Many taxes are designed to raise tax revenue, not to correct market failures. The deadweight cost associated with these taxes might be a substantial fraction of the tax revenue. Governments tax "factors of production", such as land, labor and capital, not only produced goods. Factor taxes, like commodity taxes, have an incidence and a deadweight cost. An income tax potentially alters the supply of labor, changing the equilibrium wage, affecting both people who supply labor and those who purchase it, and creating a deadweight loss.

The partial equilibrium analysis examines a single market. There, tax incidence and deadweight cost depend on the supply and demand elasticity of the commodity or the factor. If the elasticity of supply is zero, then producers or factor owners bear the entire incidence, and the deadweight loss is zero. In this case, the tax shifts income from producers or factor owners to taxpayers, but causes no efficiency loss. Unimproved land is the classic example of a factor with zero elasticity of supply. Henry George, a 19th century political economist, proposed a single tax on unimproved land; this tax causes no economic loss, and falls entirely on landowners.

A general equilibrium setting recognizes that markets for different products or factors are interlinked, and that it is seldom possible to alter one market without altering others. For example, if landowners are also farmers, the land tax lowers their income and wealth. The tax-induced reduction in income may cause farmers to work harder; collectively, the decisions change the supply of labor, thus changing the level and equilibrium price of the output. The tax-induced reduction in farmers' wealth might induce them to rebuild their wealth by accumulating more capital. The higher stock of capital increases the marginal productivity labor, increasing the equilibrium wage and lowering the return to capital. The changes in these factor prices shift the tax incidence to factor owners. If the revenue from the land tax is given to workers, their higher income might cause them to supply less labor, increasing the equilibrium wage.

These general equilibrium changes are too complicated to summarize in a simple formula. They are sometimes studied using numerical "computable general equilibrium" (CGE) models. Table 1 reports CGE-based estimates of the deadweight loss of various U.S. taxes. The property tax has the lowest deadweight loss, consistent with the low elasticity of supply of land. The fact that most taxes create a deadweight loss means that there is a social cost to raising government revenue. Prominent estimates are that this social cost is at least 20% of tax revenue. Opponents of expensive government programs sometimes invoke this cost to explain that the actual cost of the program exceeds the budgetary cost.

income tax	payroll tax	consumer sales tax	property tax	capital tax	output tax
50	38	26	18	66	21

Table 10.1 Estimates of percent deadweight loss of U.S. taxes

Time consistency Putting aside the general equilibrium complications, the partial equilibrium analysis has a simple and powerful policy message: the deadweight cost associated with taxes is lower, the lower is the elasticity of supply or demand associated with the taxed product or factor. The difference between short and long run elasticities complicates this message.

In order to make this point, suppose that the government wants to minimize the deadweight loss, subject to the constraint that it raise a certain amount of tax revenue in each period. The government has two policy instruments, a tax on capital and a tax on labor. Investment takes time; the current stock of capital depends on previous, not on current investment. This fact causes the supply of capital to be quite inelastic in the short run, so current capital taxes create little deadweight loss. This observation implies that most of the revenue in the *current* period should be raised using a capital tax. However, the future stock of capital depends on current investment, which depends on beliefs about future capital taxes. This fact causes the future stock of capital to be quite elastic, militating against the use of a capital tax in the future.

If the policymaker today can make a binding commitment to the time profile of taxes, she would like to raise most of current revenue using a capital tax, but promise to use a low future capital tax. The current labor tax can therefore be low, but the future labor tax must be high, in order to raise the required amount of revenue in each period.

The time consistency problem is that "once the future arrives, it has become the present". The policymaker in the future has the same incentives as the policymaker today. She has the temptation to renege on her predecessor's commitment to use a low capital tax, in order to raise the required amount of revenue with little deadweight cost. The planner in the future does not want to implement the program that her predecessor would like to see used, giving rise to a "time consistency problem". If today's planner cannot bind her successors to carry out the program that she wants them to use, that program is not time-consistent. Investors would not believe that they will in fact face low capital taxes in the future; this understanding reduces their incentive to invest, creating an additional distortion.

10.4 Taxes and cap & trade

Objectives and skills

- Understand the basic ingredients of a cap and trade policy.
- Understand how to use tax incidence to estimate the fraction of permits that have to be "grand-fathered" in order for a cap and trade policy not to reduce industry profits.

Emissions taxes and cap & trade are market-based environmental policies. "Command and control" policies, in contrast, reduce emissions by mandating certain types of technology or production methods. Market-based policies likely reduce emissions more cheaply than command and control policies. By the end of 2014 there were almost 50 carbon markets worldwide, the largest being the European Union's Emissions Trading Scheme (EU ETS). Southern California's trading scheme for NO_x, RECLAIM, has operated since 1994. The (never passed) 2009 Waxman-Markey bill envisioned setting up a U.S. market for carbon emissions. Environmental reform requires people to change their behavior, usually imposing a cost.

Under cap and trade, the government announces a pollution ceiling, a cap, and requires that firms have one "pollution permit" for each unit of pollution that they create; the aggregate number of permits equals the cap. The permits might be given ("grand-fathered") to polluters, or firms might be required to purchase them from the government.[2] Firms are allowed to

[2] "Grandfathering" refers to the practice of exempting certain groups from a new rule or law. The term originated during the late 19th century when Southern states created voting obstacles (such as literacy tests) to disenfranchise black citizens. White voters were exempt from these obstacles if their grandfather had voted.

trade permits; this market establishes a price for permits.

Giving (instead of selling) firms pollution permits lowers their cost of complying with the regulation. Firms facing the prospect of cap and trade want to persuade legislators to grandfather a large share of permits. Politicians may be generous, in order to mute firms' opposition to the environmental regulation. The Waxman-Markey bill proposed giving businesses a declining (over time) share of the permit allowance; economists discussed whether that plan would hand firms a windfall. Generous grand-fathering in the EU ETS may have increased polluting firms' profits. When does grand-fathering merely cushion businesses from a loss of profits, and when is it a windfall? To answer this question, we first explain the sense in which cap and trade and the pollution tax are equivalent, and then apply this insight. Our discussion uses the special case where one unit of output creates one unit of pollution. Here, a unit tax on pollution is equivalent to the same unit tax on output, so we can use the concept of tax incidence developed above.

Taxes and cap and trade are equivalent the following sense. A cap establishes a particular limit on pollution; trade in permits leads to a particular price for permits. As a pollution tax increases from zero, the equilibrium level of pollution falls. There is a "quota-equivalent" tax that results in the same level of pollution as does the particular cap. The magnitude of this quota-equivalent tax equals the equilibrium price of permits under cap and trade. Thus, the value of quota rents (= number of quotas $\times$ price per quota) equals the value of tax revenue (= amount of pollution $\times$ tax level). Grand-fathering the (arbitrary) fraction s of permits is equivalent to giving firms the fraction s of quota rents, which is equivalent to giving them the fraction s of tax revenue under the "quota-equivalent tax". Under cap and trade, rational firms' emissions decision depends on the price of an emissions permit, but not on the the number of permits they are grandfathered (Box 10.1 and Appendix F.5).

We set out to answer "What fraction of permits must be grand-fathered in order that the cap and trade policy not reduce firm profits?" The logic above shows that this question has the same answer as the question "What fraction of tax revenue would we have to give firms, in order that the 'quota-equivalent tax' not reduce firm profits?" That question has a simple answer.

Suppose that the quota-equivalent tax equals the level shown in Figure 10.1. This tax lowers firm profits by the area $abcf = abgf + bcg$. Inspection of the figure shows that the area $abgf$ equals total tax revenue ($abde$) times the producer tax incidence. Therefore, if we give firms the share of tax revenue equal to their tax incidence, plus a bit more to make up for the triangle *bcg*, firm profits are the same as before the tax. Translating this conclusion to the cap and trade context, we conclude that if firms are grandfathered a share of permits slightly greater than their tax incidence under the quota-equivalent tax, the environmental policy does not reduce firm profits.

Box 10.1 Rational firms' emissions decisions do not depend on the number of permits they are grand-fathered. If this claim is correct, and if in addition allowances were randomly assigned to firms, emissions would be uncorrelated with allowances. However, allowances are not randomly assigned: firms that emitted more in the past typically receive higher allowances. Moreover, the characteristics (e.g. old technology) that caused firms to be high emitters in the past, tend to also make them high emitters in the future. Therefore, the assignment of allowances on the basis of historical emissions creates a positive correlation between allowances and emissions. That correlation sheds no light on whether the italicized claim, above, is correct. California's RECLAIM emissions trading program randomly assigned firms to different "permit allocation cycles" which allocated allowances at different times during the year, and which tended to have different size allocations. Similar firms randomly assigned to different groups therefore tended to receive different levels of allowances. This (limited) randomness in assignment of allowances made it possible to test the italicized claim statistically; those tests support for the claim.

It is worth repeating two assumptions that underlie this conclusion. First, the economy is closed. To the extent that the policy affects electricity generators, the closed economy assumption is reasonable, because there is little international trade in electricity. However, if the policy affects producers of carbon-intensive traded goods, the results described here do not hold. If an open economy imposes cap and trade, the regulation may have little or no effect on the price at which the country can buy or sell carbon-intensive goods. Here, producers face an infinitely elastic excess demand function, and they bear the entire incidence of the regulation. These firms are worse off even if they are given all of the permits. The second assumption is that pollution is proportional to output. If changes in production methods change the ratio of emissions to output (the "emissions intensity") then the insights obtained above are still relevant, but the analysis is more complex.

10.5 Summary

The unit and ad valorem taxes are different ways of expressing a tax. For each unit tax, there is an ad valorem tax that leads to exactly the same outcome. These taxes change equilibrium price and quantity and they create a deadweight cost to society. The deadweight cost of a tax equals the net loss in social welfare arising from the tax: the reduction in the sum of consumer and producer surplus, minus the tax revenue.

The consumer incidence of the tax equals the increase in the price that consumers pay, as a percent of the tax. The producer incidence equals the reduction in the price that producers receive, as a percent of the tax. In a closed economy (no international trade) it does not matter whether the tax is imposed on consumers or producers; the two taxes are equivalent. This fact implies that in some settings, the Polluter Pays Principle is vacuous: regardless of whether the polluter or the pollutee directly pays, their actual cost is the same. The larger point is that producers and consumers typically share the burden of a regulation that increases production costs.

Consumer and producer incidences depend on supply and demand elasticities. The deadweight cost of a tax is approximately proportional to the square of the tax. Therefore, reductions in small taxes typically lead to small decreases in deadweight loss, but reductions in large taxes result in large decreases in deadweight loss. The deadweight cost might be a significant fraction of tax revenue, and can therefore significantly reduce the potential social benefit arising from tax revenue.

Three "rules" of optimal taxation state that it is better to tax commodities or factors for which the elasticity of supply or demand is small (so that the deadweight loss is small); it is better to have a broad tax base (so that a given amount of revenue can be raised using small taxes in each sector; and it is better to tax bads than goods. We are primarily interested in taxes as a means of correcting externalities such as pollution, not as a means of raising revenue. However, to the extent that taxes on bads can replace taxes on goods, the former not only correct market failures, but also potentially reduce the deadweight cost of revenue-raising taxes. General equilibrium relations can shift the tax incidence in subtle ways. In a dynamic setting, involving investment, the short run elasticities of supply and demand typically differ from their long run analogs, complicating the problem of designing tax policy, and potentially leading to a time inconsistency problem.

Under cap and trade, where firms can buy and sell emissions permits, the equilibrium price of permits depends on the level of the cap, but not on the whether the permits are given or auctioned to firms. A pollution tax equal to the equilibrium price of emissions permits leads to the same level of emissions as under a cap. If firms are given (rather than having to buy) the fraction of permits slightly greater than their tax incidence, the cap and trade policy does not lower industry profits.

10.6 Terms, study questions and exercises

New terms or concepts

Unit tax, ad valorem tax, consumer and producer tax incidence, rational buyers, behavioral economics, approximation of tax incidence, closed and open economies, Polluter Pays Principle, tax equivalence, Harberger triangle, deadweight loss (or cost) of taxes, approximation of deadweight loss, approximation of tax revenue, factor prices, general equilibrium effects, computable general equilibrium (CGE) model, time consistency, cap and trade, equivalence of a cap and trade and a tax policy, grand-fathering.

Study questions

1. (a) What does it mean to say that a producer and a consumer tax are "equivalent" in a closed economy? (Your answer should include a definition of the term "incidence". (b) Use either a graphical or a numerical example to illustrate this equivalence in a closed economy. (c) Using either a numerical or a graphical example, show that this equivalence breaks down in an open economy.

2. (a) Use a graphical example to show how the consumer and producer tax incidences (in a closed economy) depend on the relative steepness of the supply and the demand functions at the equilibrium price. (b) Using this example, explain how the approximation of consumer tax incidence depends on the demand elasticity relative to the supply elasticity (i.e. the ratio of the two elasticities) evaluated at the no-tax equilibrium. [Begin by drawing a supply demand function, picking a tax, and identifying the tax incidences. Then rotate one of the curves around the no-tax equilibrium, making it much steeper (= less elastic) or much flatter (= more elastic) and show graphically how the

incidences change.]

3. An opponent of government programs might argue that the true economic cost of financing these programs exceeds the nominal cost of the programs. An advocate of some government programs might argue that they are necessary to correct market failures. Explain, using concepts developed above, these two positions.

4. Suppose that a regulator imposes a producer tax in a closed economy. (a) Use the concept of producer tax incidence to approximate the fraction of the tax revenue that would have to be turned over to producers to make them almost as well off under the tax + transfer as they were before the tax. (b) Use the concept of consumer tax incidence to approximate the fraction of the tax revenue that would have to be turned over to consumers to make them almost as well off under the tax + transfer as they were before the tax. (c) Is it possible, by means of transferring the tax revenue (associated only with this particular tax), to make both producers and consumers exactly as well off under the tax + transfer as they were before the tax? Explain

5. (a) Describe a cap and trade policy. (Explain how it works.) (b) Explain what it means to auction permits. (c) Explain why firms' equilibrium level of pollution does not depend on whether permits are given to the firm or auctioned.

6. (a) Explain what is meant by the claim that a pollution tax and a cap and trade policy are equivalent. (b) Explain why the claim is true (in the particular setting we used). (c) Suppose that instead of using a cap and trade, a regulator uses a "cap and no trade" policy, in which firms are allocated pollution permits but not allowed to trade them. Is a pollution tax equivalent to a "cap and no trade policy"? Explain. For example, if you claim that the two policies are still equivalent, without trade, you should justify that conclusion. If you claim that the two policies are different, with respect to some significant outcome, you should identify and explain the difference.

Exercises

Assume for all questions that the economy is closed.

1. Consider a monopoly, in a static setting, with constant marginal cost c. If the monopoly receives price p and consumers pay a unit tax ν, consumers' tax-inclusive prices is $p^c = p + \nu$; if instead they pay the ad valorem tax τ, they pay the tax-inclusive price $p^c = p(1+\tau)$. The demand function is $q = a - p^c$. (a) Write down the monopoly's maximization problem and first order condition in the two scenarios, with a unit and an ad valorem tax. (b) Solve these two first order conditions to find the equilibrium consumer price in these two cases. (In one case this price is a function of ν and in the other case it is a function of τ.) (c) Use part b to find the relation between ν and τ such that if the taxes satisfy this relation, the consumer price is the same under either tax.

2. (a) Draw a linear demand function and a linear marginal cost function. Use this figure to identify (graphically) consumer and producer incidence of a unit tax, ν, in a competitive equilibrium. (b) Reproduce the figure you drew from part (a), except now make the supply function steeper at the zero-tax competitive equilibrium. Identify the consumer and producer tax incidence and compare these to your answer in part (a). (c) Explain the relation between your answer to part (b) and equation 10.3. (d) Reproduce the figure that you drew from part (a). A regulator uses a unit tax ν. Use this new figure to compare the consumer tax incidence in a competitive equilibrium and under a monopoly.

3. (a) Using the approximation of deadweight loss in equation 10.4, show that deadweight cost increases with either the elasticity of demand or the elasticity of supply. (Hint: take a derivative.) For this question, you are holding the tax and the zero-tax equilibrium quantity and price constant, and considering the effect of making either the demand or the supply function flatter (more elastic) at this equilibrium. (b) Provide an economic explanation for the relation you showed in part (a).

4. Using the assumption that the term in parenthesis in equation 10.4 is the same for each sector, show that doubling the tax base leads to a 50% reduction in the total deadweight loss (defined as the sum over the sectors of the deadweight loss in each sector).

5. A profit tax is usually expressed in ad valorem terms. If firms have profits $py - c(y)$ and pay a profit tax ϕ, their after-tax profit is $(1-\phi)[py - c(y)]$. (a) How, if at all, does a profits tax affect the equilibrium price in a competitive equilibrium? (b) What, if anything, does a profit tax affect in a competitive equilibrium? (Make a list of the features of the competitive equilibrium that we care about, and ask which if any of those features are altered by the profits tax.) (c) Now answer parts (a) and (b), replacing the competitive firm with a monopoly.

6. A monopoly has constant costs, c. Consumers, facing price p^c, demand $q = a - p^c$ units of the good; so the inverse demand is $p^c = a - q$. (a) Write down the monopoly profits, as a function of its sales, q, in the two cases where consumers pay the unit tax ν, and then when the monopoly pays the unit tax ν. (b) Compare the profit function in these two cases. Based on this comparison, does the monopoly equilibrium depend on which agent (consumers or the monopoly) has the statutory obligation to pay the tax?

Sources

Gentry (2007) reviews evidence that labor bears a significant share of the incidence of corporate taxes.

Feldstein (1977) discusses the general equilibrium effects of a land tax.

Diewert et al. (1998) and Conover (2010) review estimates of tax incidence; Table 11.1 is based on Conover.

Judd (1985) discusses the time path of capital taxes, and Karp and Lee (2003) discuss the time-inconsistency of the optimal program.

The World Bank (2014) surveys carbon markets across the world.

Fowlie, Holland and Mansur (2012) document the success of the RECLAIM market for NO_x.

Fowlie and Perloff (2013) find support for the hypothesis that emissions levels do not depend on permit allocations.

McAusland (2003 and 2008) compares environmental taxes in open and closed economies.

Sijm et al. (2006) and Hintermann (2015) provide evidence that grandfathering in the EU ETS might have given firms windfall profits.

Chapter 11

Taxes: nonrenewable resources

Objectives

- Study nonrenewable resource taxes by synthesizing the Hotelling model and facts about static taxes.

Information and skills

- Have an overview of actual fossil fuel taxes.
- Understand the time consistency problem arising from quasi-rent.
- Understand how taxes alter a firm's extraction incentives.
- Compare constant versus time-varying tax profiles.
- Map tax-induced price changes into trajectories of consumer and producer tax incidence.

Current supply in static markets depends on the current price, but in natural resource markets it also depend on firms' expectations of future prices. Some results from static tax analysis carry over to the dynamic setting: in a *closed economy* the incidence of the tax is the same regardless of whether it is levied on consumers or producers, and for every unit tax, there is an equivalent ad valorem tax. Here, there is no loss in generality in assuming that producers have the statutory obligation to pay a unit tax.

In other respects, taxes might have qualitatively different effects in a static setting and in a dynamic setting with natural resources. In a static setting, taxes reduce equilibrium supply. In the resource setting, a tax reallocates supply across periods, but possibly has no effect on *cumulative* supply. The tax-induced change in the *timing* of sales can have important welfare effects.

11.1 Current fossil fuel policies

Objectives and skills

- Know the basics of fossil fuel policies and understand the rationale for efforts to reform these policies.

Natural resources, particularly fossil fuels, are an important part of the world economy, and governments derive substantial revenue from their taxation. Between 2005 – 2010 the (mostly rich) 24 countries in the Organization for Economic Cooperation and Development (OECD) raised about $850 billion per year in petroleum taxes, including goods and services taxes and value added taxes. For large oil producing countries, government receipts from the hydrocarbon sector were a large fraction of total government revenue (2000 - 2007 data): 72% for Saudi Arabia, 48% from Venezuela, and 22% for Russia.

In many rich countries, the oil sector also receives significant implicit subsidies in the form of tax deductions. A tax on producers implicitly taxes consumption (Chapter 10); similarly, a producer subsidy implicitly subsidizes consumption. Middle income and developing fossil fuel exporters directly subsidize domestic fuel consumption by maintaining a domestic price lower than the world price. For both groups of countries, these policies subsidize fuel consumption, creating significant distortions. The fossil fuel sector also receives large implicit subsidies, because fossil fuel prices do not include the cost of externalities.

In 2009, leaders of the G20 (a group of wealthy countries) committed to "rationalize and phase out over the medium term inefficient fossil fuel subsidies that encourage wasteful consumption". A group of international organizations, including the OECD and World Bank, estimated the scope of energy subsidies and made suggestions for their reduction. Their report identified 250 individual mechanisms that support fossil fuel production in the OECD countries, having an aggregate value of USD $ 45 -75 billion per year over 2005 - 2010; 54% of this subsidy went to petroleum, 24% to coal, and 22% to natural gas. In the U.S., tax breaks provide fossil fuel subsidies of about $4 billion per year. These tax breaks include: write-offs for intangible drilling costs, a domestic manufacturing tax deduction, and a percentage depletion allowance for oil and gas wells. These subsidies transfer income from taxpayers to resource owners. Underpriced leases for mines and wells on federally owned land also transfers income from taxpayers to producers.

A group of 37 emerging and developing countries subsidize domestic fuel consumption, maintaining domestic prices below international prices. This group accounts for over half of world fossil fuel consumption in 2010. With a domestic price of p^d and a world price of p^w, the per unit subsidy is $p^w - p^d$. The nation loses $p^w - p^d$ times the amount of subsidized consumption from selling fuel at the low domestic price instead of the higher world price. Most of the countries maintained a stable domestic price, while the world price fluctuated, causing the per unit subsidy to also fluctuate. The cost of the subsidies to their domestic treasuries amounted to \$409 billion in 2010 and \$300 billion in 2009. Oil received 47% of total, and the average subsidy was 23% of the world price. The subsidy rates were highest among oil and gas exporters in Middle East, North Africa and Central Asia.

A common justification for fuel subsidies is that they benefit the poor, providing them with access to energy services. However, only 8% of the \$409 billion subsidy in 2010 went to poorest 20% of the population. If the subsidy had been eliminated, and the fuel sold at world price, and each person then given an equal share of the proceeds, the poorest 20% would have received approximately twice as much as they did under the subsidy. Fuel subsidies – like most commodity subsidies – are an inefficient way to help the poor. These subsidies fell from about 1.8% of government budgets in 2004 to 1.3% in 2010. Absent reforms, 2011 estimates project that these fossil fuel subsidies would reach \$660 billion per year by 2020. The elimination of consumption subsidies was estimated to reduce 2020 fuel demand by 4.1% and CO_2 emissions by 4.7%.

A 2015 International Monetary Fund (IMF) study updates estimates of the magnitude and economic cost of fossil fuel subsidies, and also includes unpriced externalities. For fossil fuels, the implicit subsidy arising from the unpriced externality is larger than the direct subsidy. Local health effects, not climate change, accounts for the bulk of this externality cost. The IMF study estimates that global energy subsidies (including the externality cost) amounted to about \$5 *trillion*, or 6% of world Gross Domestic Product (GDP) in 2013; removal of these explicit and implicit subsidies would have raised almost \$3 trillion in government revenue, and would have increased global GDP by more than 2%. For comparison, estimates of the increase in welfare due to major trade liberalization are typically only a fraction of a percent of GDP. A 2014 IMF book estimates that efficient energy prices (eliminating explicit subsidies and imposing externality costs) would reduce carbon emissions by 23% and reduce deaths related to fossil-fuel air pollution by 63%.

Fossil fuel subsidies are inefficient for at least four reasons. Most importantly, they subsidize a commodity that should, because of environmental externalities, be taxed. The subsidies also violate the other two "rules" of optimal taxation discussed in Chapter 10.3. The first rule states that governments should have a broad tax base, so that taxes on each sector can be low. Subsidizing rather than taxing the fossil fuel sector flies in the face of this advice, requiring higher taxes in other sectors to finance the fossil fuel subsidies. The second rule states that, for the purpose of raising government revenue, goods with inelastic supply and/or demand should be the most highly taxed. Fossil fuels have inelastic short run supply and demand, but they are subsidized. Finally, commodity subsidies are an inefficient means of making transfers to the poor. Political power, not economic logic, explains the tax and subsidy policies used in a wide range of fossil fuel markets, for both importers and exporters countries, and for rich and developing nations.

Renewable energy sources also receive significant subsidies. The dollar value of these is much smaller than the value of "direct" fossil fuel subsidies (i.e., excluding the unpriced externalities associated with fossil fuels). However, renewables account for a small part of the energy market. The (direct) subsidy per unit of energy produced is 2 –3 times larger for renewables than for fossil fuels. This ratio overstates renewables' subsidy advantage, relative to fossil fuels, because the renewable subsidies have fluctuated over the past decades, creating a risky investment climate; fossil fuel subsidies have been maintained by political influence. Fossil fuels also rely on relatively mature technologies, compared to renewables; positive externalities such as learning-by-doing and research spillovers (standard rationales for subsidies) are consequently more plausible for renewables than for fossil fuels.

11.2 The logic of resource taxes

Objectives and skills

- Understand the effect of taxes on the timing of extraction, and the potential effect of taxes on cumulative extraction.

- Understand how taxes affect incentives to develop resource stocks.
- Understand governments' temptation to raise resource taxes after firms have made expensive investments, and the resulting "hold-up" problem.

In the static setting, taxes (and subsidies) drive a wedge between consumer and producer prices, altering equilibrium quantity and creating deadweight loss. Natural resource taxes can alter both the timing of extraction and cumulative extraction. We focus on the timing effect by studying a model with constant average extraction cost C, where taxes do not alter cumulative extraction. We then discuss the relation between taxes and investment.

The Euler equation provides the basis for understanding how taxes alter sales and price trajectories. This equation requires the present value of rent to be constant over time. Equation 5.6, repeated here, is

$$R_t = \rho^j R_{t+j}. \tag{11.1}$$

In the absence of tax, rent is $R_t = p_t - C$. Under the tax $\nu(t)$, if the consumer price is p_t, producers' (after tax) price is $p_t - \nu(t)$, so their rent is $R_t = p_t - \nu(t) - C$. With this revised definition of rent, the firm's optimality condition is still equation 11.1, or

$$p_t - \nu(t) - C = \rho^j \left(p_{t+j} - \nu(t+j) - C\right). \tag{11.2}$$

A thought experiment helps in understanding the effect of taxes on the equilibrium price trajectory. Suppose that we begin with the no-tax equilibrium where the (producer = consumer) prices, p_t^{NT}, (NT for "no-tax") satisfy the Euler equation:

$$p_t^{NT} - C = \rho^j \left(p_{t+j}^{NT} - C\right). \tag{11.3}$$

We now impose a tax sequence, $\nu(t)$, $t = 0, 1, 2...$ and ask whether the no-tax prices still constitute an equilibrium. We write this question as

$$p_t^{NT} - \nu(t) - C \stackrel{??}{=} \rho^j \left(p_{t+j}^{NT} - \nu(t+j) - C\right). \tag{11.4}$$

The symbol "$\stackrel{??}{=}$" indicates that we are asking whether the equality holds; if it does not hold, we want to determine the change in the price trajectory that makes it hold. Two examples show how the thought experiment provides information about the effect of resource taxes. Here, unlike the static setting (equation 10.3), we do not have simple formulae for measuring tax incidence. However, the resource firm's equilibrium condition tells provides information about tax incidence without performing calculations.

A tax that increases at the rate of interest

First consider $\nu(t) = \nu_0 (1+r)^t$, a tax increases at the rate of interest. For this tax,

$$\nu(t+j) = \nu_0 (1+r)^{t+j} = \nu_0 (1+r)^t (1+r)^j = \nu(t)(1+r)^j$$
$$\Rightarrow \rho^j \nu(t+j) = \nu(t).$$

We obtain the second line by multiplying the first line through by ρ^j, and using $\rho^j (1+r)^j = 1$. *If the tax increases at the rate of interest, then the present value of the tax is constant.* Subtracting $\nu(t)$ from the left side of equation 11.3 and the same quantity, $\rho^j \nu(t+j)$, from the right side, we obtain equation 11.4 (replacing $\stackrel{??}{=}$ with $=$). We conclude that if the present value of the tax is constant (as assumed here), it has no effect on equilibrium consumer prices: the consumer incidence is 0%. Therefore, the producer incidence must be 100% (because the two incidences sum to 100%). Because the tax does not alter the equilibrium consumer price, it does not alter the equilibrium sales trajectory, and it creates no deadweight loss.

The present value of the tax receipts is

$$\sum_{t=0}^{T} \rho^t (1+r)^t \nu_0 y_t = \nu_0 \sum_{t=0}^{T} y_t = \nu_0 x_0$$

This tax transfers the rent $\nu_0 x_0$ from producers to taxpayers, without creating a distortion. Equation 5.8 shows that under constant extraction costs and zero tax, the value of the firm is $R_0^{NT} x_0$, where R_0^{NT} is the No Tax initial rent. Therefore, under the tax considered here, the after-tax value of the firm is $\left(R_0^{NT} - \nu_0\right) x_0$. By setting the initial tax, ν_0, close to R_0^{NT}, the government can extract nearly all of the rent from the resource owner.

A constant tax

Suppose now that firms face the constant unit tax, ν. In moving from equation 11.3 to 11.4 (with constant ν), we subtracted ν from the left side and $\rho^j \nu$ from the right side. Because $\nu > \rho^j \nu$, the "questioned equality" in 11.4 is false. In order for the equality to hold, we have to increase p_t relative to p_{t+j} (because we are subtracting a larger quantity from the left than from the right side). We can make this adjustment by transferring sales from period t to period $t+j$, thereby increasing the price in period t and reducing it in period $t+j$.[1]

The conclusion is that the constant tax reduces sales and increases the price early in the program, and has the opposite effect late in the program. Therefore, the consumer incidence of the tax is positive early in the program (when the tax leads to a higher consumer price), but it is negative late in the program (when the tax leads to a lower consumer price). This example shows that the consumer tax incidence need not lie between 0% and 100% in the resource setting, unlike in the static setting.

The constant tax has the same qualitative effect as a higher cost, discussed in Chapter 3.3. There, we saw that a higher extraction cost causes firms to shift production from the current period to future periods. Discounting reduces the present value of costs that are incurred in the future. From the standpoint of the firm, increasing the tax from 0 to the positive constant ν has exactly the same effect as increasing costs from C to $C' = C + \nu$.

In a static model, competitive firms with constant marginal production costs have infinite elasticity of supply. Here, the consumer tax incidence is 100% and the producer incidence is zero (equation 10.3). In the resource setting, the consumer tax incidence is positive early in the program but negative late in the program. What explains the difference (under constant marginal costs) between the static and resource settings? In the resource setting, total supply is finite; marginal extraction cost is constant before exhaustion, but infinite once the resource is exhausted. The resource scarcity creates rent. In contrast, in the static setting, producer surplus is zero.

Box 11.1 The profits tax Exercise 2 in Chapter 10 explores the effect of a constant profits tax in a static model. That exercise shows that a constant profits tax takes rent from producers, but has no effect on equilibrium price or quantity. This result also holds for resource markets. The constant profits tax reduces profits in each period proportionally, and does not alter the firm's incentives about when to produce. A time varying profits tax alters the firm's incentives, and therefore affects consumers in addition to firms.

[1] It is worth repeating the warning made in Chapter 3.1. The firm does not respond to an exogenous change, such as a new tax, *in order* to cause equation 11.2 to hold. The exogenous change presents the price-taking resource owner with new opportunities for intertemporal arbitrage. The firms' response causes prices to change. Additional opportunities for intertemporal arbitrage are exhausted only when equation 11.2 holds.

Investment

Rent is the return to a factor of production in fixed supply; quasi-rent is the return to a previous, now "sunk", investment. Most natural resource stocks require significant investment in exploration and development. Thus, the difference between the price that a firm receives and its marginal cost of extracting the resource, is actually the sum of rent and quasi-rent. We referred to this sum as "rent" merely in the interest of simplicity. For the purpose of studying tax policy, the distinction is important.

Suppose that the initial rent + quasi-rent for a mine with constant extraction costs is $R_0 = 10$ and the initial stock is $x_0 = 10$, so the value of this mine (using equation 5.8) is 100. Each mine costs 5 to develop, and on average a firm must develop five mines to find one that is successful. The firm can test many potential mines at the same time, understanding that on average 20% of them will be successful. It is important to account for all of the unsuccessful mines in estimating the development cost of a successful mine. In this example, the expected investment cost for a successful mine is 25. Here, the rent on the mine is 75 and the quasi-rent is 25.

Prior to the exploration and development, the government might announce a tax that begins, at the time of initial extraction, at 7.5, and *rises at the rate of interest*. With this tax, the government captures all of the rent, but leaves firms with the quasi-rent. In this case, firms break even. (As a practical matter, the tax should leave firms with some rent, in order to induce them to undertake the risk; we ignore that complication, in assuming that firms are risk neutral.) Once a successful mine is in operation, the government might be tempted to renege on its announcement, and to impose an initial tax of $\nu_0 = 10$, which increases at the rate of interest. This tax transfers all of the rent + quasi-rent from the firm to taxpayers. The firm in this case loses all of its sunk investment costs, and suffers a net loss of 25. This temptation to renege illustrates the potential time-inconsistency of optimal plans. The situation where one agent (here the government) takes advantage of a second agent's (here the resource firm) sunk investment is known as a "hold-up problem".[2]

The real-world importance of the hold-up problem varies with the setting. If there are many cycles of investment and extraction, then the government

[2] "Hold up" has two possible meanings here, either as a robbery or as a delay (because the first agent's incentive to take advantage causes the second agent to delay in undertaking the investment).

has an incentive to adhere to its promises in order to maintain its reputation. By behaving opportunistically, this government obtains short run benefits, in the form of higher tax revenue, but it discourages future investment. This hold up therefore tends to be more important for a single large project, e.g. a one-time development of offshore oil deposits. (These considerations also apply to non-resource related infrastructure projects, e.g. building a harbor or developing a transportation network.) The government's temptation to renege may be greater if the major investors are foreigners, because domestic investors might be better able to defend their interests in the political arena. For large one-off foreign-sourced investment projects, hold up can be a major issue. Without some protection, foreign investors would be unwilling to undertake the project, or would require a large risk premium for doing so.

The OECD and other international organizations attempted, during the 1980s and 1990s, to negotiate a Multilateral Agreement on Investment to resolve this hold-up problem. This agreement was advertised as a means of encouraging international investment, in much the same way that the World Trade Organization promotes international trade. Facing resistance from developing countries, the Multilateral Agreement on Investment was never completed. However, there are currently thousands of Bilateral Investment Treaties (BITs), most of which involve one rich and one developing country; the U.S. is party to over 40 BITs.

The parties of these treaties are countries, not private firms, but many of the treaties have an "investor-to-state" provision. This provision permits a private *investor* originating in one signatory (usually the rich country) to sue, in an international court, the *government* of the other signatory (usually the developing country) for violation of the treaty. A primary purpose of the treaties is to protect an investor against confiscation of their investment, but many treaties also provide protection against measures that are "tantamount to expropriation", such as "confiscatory taxes" or even post-investment changes to environmental rules. Business interests regard the investor-to-state provision of these treaties as essential, because they are not confident that their own government would act in their interests. For example, foreign policy considerations might make the U.S. government reluctant to invoke a treaty in protection of a U.S. investor. The investor, in contrast, has no such qualms about exercising the treaty rights.

The treaties provide a self-enforcing mechanism, a credible commitment against opportunistic behavior. However, they also limit a country's ability to exercise what are known as "police powers" (e.g. environmental regulation) and to respond to contingencies not foreseen at the time of investment (e.g. a major recession). Although business groups strongly support these treaties, some NGOs think that they harm developing country interests. Investor-to-state provisions were a major source of controversy in the North American Free Trade Agreement (NAFTA) and they have been a reason that some NGOs oppose the Trans-Pacific Partnership.

11.3 An example

Objectives and skills

- Based on a numerical example and graphs, understand: (a) the effect of the "shape" of the tax trajectory on equilibrium price trajectories, and the resulting trajectories of tax incidence; (b) the welfare changes associated with a tax trajectory; (c) the difference between anticipated and unanticipated taxes.

A numerical example illustrates the effect on resource markets of a time-varying tax, $\nu(t) = (t+1)^{\kappa}$, with $\nu(0) = 1$. Figure 11.1 shows that larger values of κ imply a more rapid increase in the tax. The effect of the tax on the incentive to extract is intuitive. If κ is close to 0, the tax grows slowly, and the situation is similar to the constant tax case discussed above. Here, the tax creates an incentive to delay extraction, as a means of decreasing the present value of the tax liability. If κ is large, the tax grows quickly. Here, the firm has an incentive to accelerate extraction so that more of its sales incur the relatively low current taxes instead of high future taxes, reducing the firm's present value tax liability.

11.3.1 The price trajectories

We use the steps discussed in Chapter 5.4 to calculate the equilibrium consumer and producer price trajectories under a particular tax trajectory, with demand $= 10 - p$, constant extraction costs $C = 1$, initial stock $x_0 = 20$, and discount rate $r = 0.04$. For the zero tax, $\nu = 0$, the resource is exhausted at $T = 11.2$. For the slowly growing tax, the terminal time is only slightly greater, $T = 11.5$. The rapidly growing tax gives producers a strong incentive to extract early, while the tax is still relatively low, leading to a much earlier exhaustion time, $T = 7.5$. Figure 11.2 shows the equilibrium price

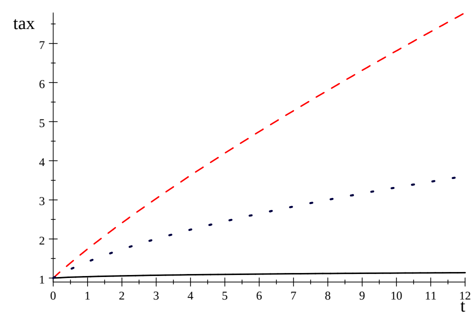

Figure 11.1: The tax profile for $\kappa = 0.05$ (solid), $\kappa = 0.5$ (dotted) and $\kappa = 0.8$ (dashed).

trajectories under these three tax profiles. Each trajectory reaches the choke price, $p = 10$, at the time of exhaustion.

The equilibrium price trajectories under the zero tax (dotted) and under the slowly increasing tax (solid) are almost indistinguishable. The positive tax encourages firms to delay extraction, increasing the initial price; the increasing tax profile encourages firms to move extraction forward in time, reducing the initial price. These two effects almost cancel. In contrast, the steeply rising tax trajectory gives firms a much stronger incentive to extract early, while the tax is still relatively low. The steeply rising tax therefore leads to a substantially lower initial price and earlier exhaustion. (Compare the dashed and the solid graphs.)

11.3.2 Tax incidence

As in Chapter 10.1, the consumer incidence is defined as the difference in the consumer price with and without the tax, divided by the tax, times 100 (to convert to a percent). We obtain the producer incidence by subtracting the consumer incidence from 100. The tax causes a reallocation of supply over time, but no change in cumulative supply (equal to the initial stock) over the life of the resource. Thus, for some periods the tax lowers the equilibrium consumer price, leading to negative consumer tax incidence and producer tax incidence above 100%. In a static competitive model both the consumer and producer tax incidences lie between 0 and 100%.

Figure 11.3 graphs the producer and the consumer tax incidences under the slowly increasing tax. The consumer incidence begins at about 15% and

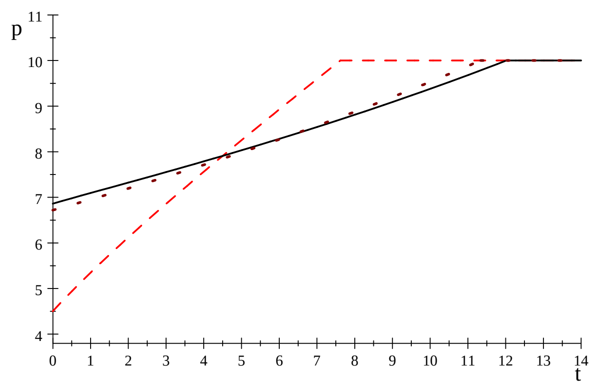

Figure 11.2: The equilibrium consumer price trajectory under the tax $\nu = (t+1)^{0.05}$ (solid); under the tax $\nu = (t+1)^{0.8}$ (dashed); and under zero tax, $\nu = 0$ (dotted).

becomes negative at the time when the dotted and solid curves in Figure 11.2 cross, $t = 6.5$. At later dates, the tax reduces the equilibrium consumer price, so the consumer incidence is negative there, and the producer incidence exceeds 100%. Figure 11.4 graphs the consumer and producer tax incidence over time under the rapidly growing tax. This tax lowers the equilibrium consumer price for $t < 4.5$, so over this region, the consumer incidence is negative, and the producer incidence exceeds 100%; at later dates, the tax incidence for both consumers and producers lies between 0 and 100%.

11.3.3 Welfare changes

Here we consider the resource tax's welfare cost. If the consumer price in period t is p, consumer surplus $(CS\,(t))$, producer profit $(PS\,(t))$, and tax revenue ($G\,(t)$, for "government") in that period equal, respectively,

$$\begin{array}{c} CS = \int_p^{10} (10-q)\,dq = 50 - 10p + \frac{1}{2}p^2 \\ PS = (p - \nu\,(t) - 1)\,(10-p)\,, \text{ and } G = \nu\,(t)\,(10-p)\,. \end{array} \tag{11.5}$$

The equilibrium p changes over time, so the functions CS, PS, G also change over time. We define an agent's welfare as the present discounted value of their stream of single period payoffs. Welfare for consumers, producers, and

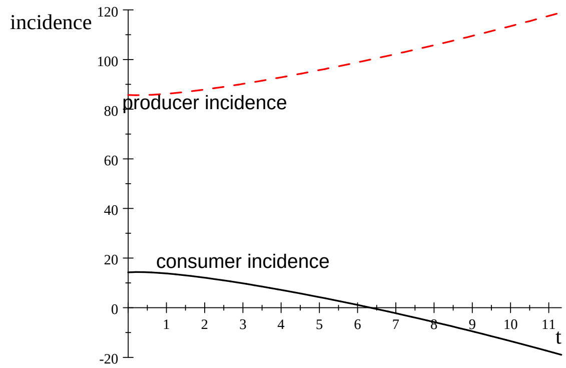

Figure 11.3: Consumer and producer tax incidence under the slowly increasing tax

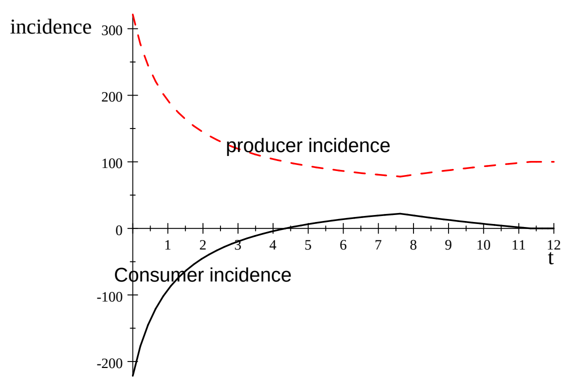

Figure 11.4: Consumer and producer tax incidence under the rapidly increasing tax

taxpayers equal, respectively,

$$\int_0^T \rho^t CS(t)\,dt, \int_0^T \rho^t PS(t)\,dt, \int_0^T \rho^t G(t)\,dt. \tag{11.6}$$

The sum of these three integrals equals social welfare. Table 11.1 shows welfare for consumers, producers, and taxpayers, and the sum of these three welfare measures (social welfare) under the zero tax and for both the slowly and the rapidly growing tax. The table also shows the percent change in welfare for consumers, producers, and society as a whole, in moving from the 0 tax to either of the two positive taxes.

	consumer welfare	producer welfare	taxpayer welfare	social welfare	$\frac{\text{DWL}}{\text{tax rev}}\times 100\%$
zero tax	20.2	114.4	0	134.6	
"slow tax"	19	97.2	18.3	134.5	0.55%
% change	-5.9%	-15%	NA	-0.07%	
"fast tax"	33.1	50.1	47	130.2	9.4%
% change	+64%	-56.2%	NA	-3.3%	

Table 11.1: Agents' welfare and % change in welfare under the zero tax, the "slow tax" $\nu(t) = (t+1)^{0.05}$, and the "fast tax" $\nu(t) = (t+1)^{0.8}$. "% change" is relative to $\nu = 0$.

Just as in the static model, the slowly growing tax reduces consumer and producer welfare, here by 6% and 15%, respectively. The increase in tax revenue almost offsets those two reductions, so social welfare under this tax falls by less than a tenth of 1%. Social welfare falls by over 3% under the rapidly growing tax. In discussing Figure 10.3 for the static model, we noted that taxes create deadweight loss that is proportional to the square of tax. The deadweight loss is negligible at a small tax, but it increases faster than the tax. The last column shows the social loss as a percent of tax revenue (both expressed as present discounted sums); for the slowly growing tax, the loss is about half a percent of tax revenue, and for the rapidly growing tax it is about 9% of tax revenue.

In the static setting, the tax increases the equilibrium price that consumers face, and thus lowers consumer welfare. In contrast, in the nonrenewable resource setting, the tax increases price in some periods and lowers prices in other periods. Depending on the magnitude and the timing of the changes, consumer welfare might either increase or decrease. Figure 11.2 shows that the rapidly growing tax reduces consumer price early in the trajectory, and increases it late in the trajectory. The tax thus increases consumer surplus early in the trajectory and lowers it late in the trajectory. Because of discounting, the early increases count for more. Thus, the rapidly growing tax increases the present discounted value of the stream of consumer surplus. This tax benefits consumers and leads to a large increase in tax revenue, but it also causes a large decrease in producer welfare, resulting in a 3.3% fall in aggregate social welfare.

11.3.4 Welfare in a dynamic setting

We proceeded as if using the present discounted stream of surplus to measure consumers' and producers' welfare is so obvious as not to require comment. Although this procedure is standard in economics, it has a shortcoming that we mention here, and take it up more carefully in Chapter 19. The resource consumers alive today are not the same people who will live in 100 years. What does it mean to say that "consumer welfare" increases, when one group of consumers is better off and another worse off? The ethical objection to our welfare measure is that it adds up the utility of people who are alive at different points in time. Moreover, it does so in a way that privileges those currently living, because the welfare measure discounts the surplus of future generations of consumers.

Our measure of social surplus at a point in time also adds up the surplus of possibly different people, consumers and producers. That aggregation is ethically less questionable, for two reasons. First, consumers and producers may or may not be different people, whereas individuals living today and in 100 years are certainly different people. Second, consumers and producers currently living have at least the potential to influence current government policy that affects their well-being. People living in the future have no direct voice in the current political process.

11.3.5 Anticipated versus unanticipated taxes

The examples above consider scenarios when a tax is introduced at the beginning of the problem, at time 0. Here we compare the no-tax case with two scenarios where a constant tax, $\nu = 3$, is introduced at a future time, $t = 3$. When the tax is unanticipated, producers have made no provisions for it. In contrast, producers who anticipate the tax adjust their sales even before the tax is imposed. Our example uses the demand function $q = 10 - p$, with constant costs $C = 1$ and discount rate $r = 0.03$. Figure 11.5 shows the trajectories of consumer (tax-inclusive) price in the three scenarios. Figure 11.6 shows the trajectories of rent (price – cost – tax) in the three scenarios.

By definition, an unanticipated tax cannot alter anything before the tax begins. Therefore the price and rent trajectories, under the unanticipated tax, are coincident with the 0-tax trajectories prior to $t = 3$. (The solid curves cover the dashed curves for $t < 3$.) At $t = 3$, when producers discover that they will begin to face the tax, the tax-inclusive cost of providing the commodity suddenly increases, causing producers to lower sales relative to the no-tax scenario. The equilibrium consumer price jumps up and the rent jumps down.

In the case of an anticipated tax (dotted curves), producers have an incentive to sell a larger amount (relative to the no-tax scenario) early in the program, before the tax begins. These high sales lead to a low initial price trajectory. The price jumps up at $t = 3$, when the tax begins. The trajectory for rent is continuous (dotted curve in Figure 11.6) under the anticipated tax, rising at the rate r. In the absence of surprises, the equilibrium rent takes into account ("capitalizes") future changes (here, the change in the tax from 0 to a positive level).

Denote the equilibrium price in the period (or instant) before the tax increase as p^- and the price immediately after the tax as p^+, so the tax causes the price to jump at t by $\Delta \equiv p^+ - p^-$. For two reasons, the jump, Δ, is larger in the case of the anticipated tax compared to the unanticipated tax. First, anticipation of the tax led to a lower pre-tax price: p^- is lower under the anticipated tax, as seen by comparing the dotted and the dashed/solid curves in Figure 11.5 at $t < 3$. Second, the lower price under the anticipated tax led to higher extraction, leaving less stock in that scenario (relative to the unanticipated case). After time $t = 3$, producers in both scenarios expect to face the tax. However, the producer in the anticipated tax scenario has (at time $t = 3$) a smaller stock compared to the producer in the unanticipated tax scenario. The lower $t = 3$ stock in the former case leads to higher prices; compare the dotted and the dashed curves in Figure 11.5 for $t > 3$.

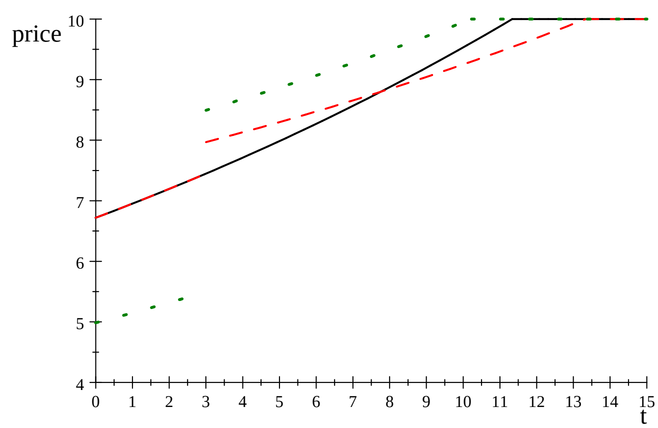

Figure 11.5: Solid curve shows the trajectory of consumer price in the absence of a tax. Dashed trajectory: consumer price when the tax imposed at $t = 3$ is *unanticipated*. Dotted trajectory: consumer price when the tax at $t = 3$ is *anticipated* at $t = 0$.

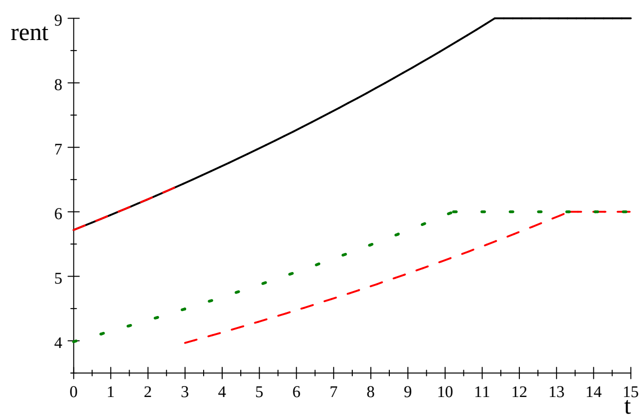

Figure 11.6: Solid curve shows the trajectory of rent in the absence of a tax. Dashed trajectory: rent when the tax imposed at $t = 3$ is *unanticipated*. Dotted trajectory: rent when the tax at $t = 3$ is *anticipated* at $t = 0$.

11.4 Summary

The theory of optimal taxation suggests that fossil fuels should be relatively heavily taxed, in order to offset the negative externality associated with their consumption, and to take advantage of their relatively low elasticities of supply and demand. Both rich and developing nations, and both importers and exporters, subsidize consumption of fossil fuels. Political constraints impede international attempts to move toward more rational resource policies.

Although taxes have more complicated effects in nonrenewable resource markets compared to static markets, the methods developed in earlier chapters make the outline of the analysis straightforward. A constant tax on extraction encourages firms to delay extraction, raising consumer prices early in the program and lowering prices later in the program. A tax that increases at the rate of interest has no effect on the extraction trajectory, because the present value of this tax is constant. A tax that increases rapidly over time increases firms' incentive to extract early, before the tax is high. This effect tends to offset the incentive to delay extraction. If the tax increases very slowly, then the "delay incentive" is the stronger of the two. However, if the tax increases rapidly, producers respond to the tax by moving the extraction profile forward in time. Thus, a rapidly increasing tax lowers price early in the program and increases later prices.

In general, taxes lead to a reallocation of supply over time. Therefore, there are some periods when the tax increases supply, relative to the no-tax case; for those periods, the consumer incidence of the tax is negative, and the producer incidence of the tax exceeds 100%. These periods of negative consumer tax incidence tend to occur late in the program, if the tax is constant or increases slowly; those periods tend to occur early in the period if the tax increases rapidly.

The zero-tax competitive equilibrium maximizes social surplus, so in the setting here a tax of any nature reduces social surplus. In the static competitive setting, both consumers and producers bear some of the incidence of the tax. The tax therefore decreases the welfare of both agents. In the resource setting with stock-independent extraction costs, the tax shifts production from one period to another, without altering cumulative production. In this case, the tax must lower price and therefore increase consumer surplus in some periods. The tax might either increase or decrease the present discounted value of the stream of consumer surplus. However, the tax reduces the producer price (equal to the price consumers pay minus the unit tax) in every period. Therefore the producer incidence is positive in every period. In addition, the tax lowers producer profit in every period. Therefore, the tax necessarily lowers the present discounted stream of producer profit.

We compared the effect of anticipated versus unanticipated taxes. The effect of an anticipated tax is capitalized into the resource price even before the tax comes into effect. Therefore, the anticipated tax alters the equilibrium even before it starts. By definition, an unanticipated tax cannot affect anything before it begins. Therefore, the implementation of an unanticipated tax tends to create a larger change in equilibrium price, at the time it begins, relative to an anticipated tax.

We emphasize the relation between taxes and extraction decisions, usually keeping the investment decision in the background. However, investment is important in resource markets, creating "quasi-rents", the return to a previous investment. Price minus marginal costs, which we usually refer to as "resource rent" is in fact the sum of genuine rent and quasi-rent. The fact that investment is sunk at the time of extraction creates a time consistency or holdup problem for policy makers. This problem is likely most severe in the case of large, one-off foreign investments. Bilateral investment treaties, with an investor-to-state provision, attempt to solve this hold-up problem.

11.5 Terms, study questions, and exercises

Terms and concepts

Tax trajectories, hold-up problem, rate of change of a tax (or of anything else), trajectories of producer and consumer tax incidence, intertemporal welfare (the integral, or the sum, of the discounted stream of consumer or producer welfare).

Study questions

For all these questions, assume that the average = marginal extraction cost is constant with respect to extraction and independent of the stock of the resource.

1. (a) Explain how a *constant* tax alters the competitive nonrenewable resource owner's incentives to extract, and thus how the tax affects the equilibrium price trajectory. (b) Now explain how an *increasing*

tax trajectory alters the competitive resource owner's incentives, and thereby alters the equilibrium price trajectory. (Your answer to both parts should make clear how the taxes affect the relative advantage of extracting at one point instead of another.) (c) Use your answers to parts (a) and (b) to explain why the three tax profiles shown in Figure 11.1 give rise to the three price profiles shown in Figure 11.2.

2. Using Figure 11.2, sketch the consumer and producer tax incidence over time, for the slowly growing and the rapidly growing tax. (Figures in the text actually show those graphs. You should see whether you can produce the sketches using only Figure 11.2 and then compare your answers with the figures in the text. Your sketches will not get the magnitudes correct, but they should correctly show the intervals of time where an incidence is negative, positive and less than 100%, or greater than 100%. If you do it carefully, you can also see where the incidences are increasing or decreasing over time. The point of this question is to see whether you REALLY know what tax incidence means.

3. (a) Explain why, in the static setting, a tax always reduces consumer surplus. (b) Explain why, in the resource setting, the tax must increase consumer surplus *at some points in time*. (c) Explain why your answer to part (b) implies that a tax might either increase of decrease the present discount stream of consumer surplus. (d) Is there any objection to using the present discounted stream of consumer surplus as a measure of consumer welfare?

Exercises

1. Suppose that the government uses a profits tax, $\phi(t)$, instead of a unit tax. This profits tax equals $\phi(t) = 1 - \exp(-\kappa t)$ with $\kappa > 0$. (a) Sketch two graphs of this tax, as function of time, on the same figure, for a small and a large value of κ. (b) Following the logic in the text for the increasing unit tax, briefly explain the effect of an increasing profits tax on the equilibrium extraction profile. (You have to think about how the increasing profits tax affects the firm's incentive to extract the resource. (c) Compare the equilibrium effect (on the extraction profile) of a *constant* unit tax and a *constant* profits tax. What explains this difference between the two types of constant taxes?

2. You need functional forms and parameter values to calculate the tax-incidence graphs shown in Figures 11.3 and 11.4, but you can infer their general shape by inspection of Figure 11.2, and from the definition of "tax incidence". In a couple of sentences (using the definition of tax incidence) explain how to make this inference.

3. The text considers the effect of a tax in a model with stock-independent extraction costs. How might stock dependent extraction costs alter the trajectories of tax incidence. (This question calls for intelligent speculation, not calculation.)

4. Suppose that firms have constant extraction costs and face a time varying profits tax, $\phi(t) = (0.05)\gamma^t$. Under this tax, if extraction in period t is y, a firm's after-tax profit is $(1 - \phi(t))(p(y)y - C)$. (a) On the same figure, graph $\phi(t)$ (as a function of t) for $\gamma < 1$ and also for $\gamma > 1$. (b) Using intelligent speculation (not calculation), explain how these two profits tax affects the equilibrium price and extraction paths. (Figures will make your explanations clearer.) Provide the economic logic for your answer.

5. Justify the claim that the functions in equation 11.5 do indeed equal consumer and producer welfare and tax revenue.

Sources

Sinclair (1992) points out the effect of a rising carbon tax on the incentive to extract fossil fuels.

Boadway and Keen (2009) survey the theory of resource taxation.

Chapter 11.1 is based largely on the IEA, OPEC, OECD and World Bank Joint Report (2011).

The IMF paper by Coady et al (2015) provides estimates of the cost (including environmental costs) of resource subsidies.

Parry et al (2014) illustrates the design of efficient energy taxes for 150 countries.

Aldy (2013) discusses the fiscal implications of eliminating U.S. fossil fuel subsidies.

Daubanes and Andrade de Sa (2014) consider the role of resource taxation when the discovery and development of new deposits is costly.

Lund (2009) reviews the literature on resource taxation under uncertainty.

Bohn and Deacon (2000) discuss investment risk with natural resources.

Aisbett et al. (2010) discuss investment risk and bilateral investment treaties.

Chapter 12

Property rights and regulation

Objectives

- Understand how property rights alter the problem of second-best regulation.

Information and skills

- Have an overview of the consequences of and the evolution of property rights.
- Be familiar with the Coase Theorem, and understand its relevance to policy in the presence of externalities.
- Have an overview of fishery regulation and subsides, and understand how these can lead to overcapitalized fisheries.
- Understand effects of individual quotas on property rights and resource outcomes.

We have emphasized competitive equilibria in nonrenewable resource markets (e.g., oil, coal) with perfect property rights. This chapter sets the stage for a discussion of renewable resources (e.g. fish, groundwater, forests, the climate), emphasizing the role of imperfect property rights. We begin with a general discussion of property rights and then show that the emergence of an efficient outcome depends on the existence but not on the *allocation* of property rights (i.e., who possesses the property rights). This result is known as the Coase Theorem.

We then discuss the role of property rights and regulation in fisheries. Fisheries provide a natural focus for this discussion, because: they are economically important; they are plagued by imperfect property rights; there is a large body of research devoted to their study; and the insight gained from studying fisheries is applicable to many other resource problems. Due to overfishing, loss of habitat, and climate change, 30% of the world's fisheries are at risk of population collapse. Fisheries support nations' well-being through direct employment in fishing, processing, and services amounting to hundreds of billions of dollars annually. Fish provide nearly 3 billion people with 15 percent of their animal protein needs, helping support nearly 8% of the world's population.

12.1 Overview of property rights

Objectives and skills

- Know the characteristics of the three types of property rights.

A spectrum of social arrangements govern the use of natural resources. The three leading modes are private property, common property, and open access. With private property, an individual or a well-defined group of individuals (e.g. a company) owns the asset and determines how it is used. Under common property, use of the asset is limited to a certain group of people, e.g. those living in a town or an area; community members pursuing their individual self-interest, instead of a single agent, decide how to use the asset. Anyone is free to use an open access resource. A farm owned by a person or corporation is private property. A field on which anyone in the village can graze their cows, but from which those outside the village are excluded, is common property. A field that anyone can use is open access.

This taxonomy identifies different types of ownership structure, but in practice the boundaries between them are often blurred. Labor, health, and environmental laws govern working conditions and pesticide use on privately owned land, limiting the exercise of private property rights. In Britain, common law allows anyone to use paths across privately owned farmland, provided that they do not create a nuisance, such as leaving gates open, further limiting private property rights; in Norway, people are allowed to enter uncultivated private property to pick berries. Common property dilutes property rights, but social norms often limit community members' actions.

Anyone in the village may be allowed to use the village common, but they might be restricted to grazing one cow, not ten. Even for an asset that is nominally open access, members of a group might exert pressure to restrict outsiders. Local surfers at some California public beaches make it uncomfortable or dangerous for outsiders to surf.

There is a continuum of types of property rights, not three neat types. *De jure* property rights describe the legal status of property, and *de facto* property rights describe the actual property rights. In the surfing example, the *de jure* property rights are open access, but to the extent that the local surfers are successful in excluding outsiders, the *de facto* property rights more closely resemble common property.

Property rights to a resource may change over time, often responding to migration and increased trade, and often accompanied by social upheaval.

- The enclosure movement in the UK, converting village commons to private property, began in the 13th century and was formalized by acts of parliament in the 18th and 19th century. These enclosures increased agricultural productivity, dispossessing rural populations.

- In the late 19th and early 20th century Igbo groups in Nigeria converted palm trees from private to common property in response to increased palm oil trade. Trade increased the value of the palm trees, increasing the need to protect them from over-harvesting. In this case (but not in general) monitoring and enforcement costs needed to protect the resource were lower under common property.

- The 1924 White Act in Alaska, later incorporated into the state's constitution, abrogated aboriginal community rights to the salmon fishery. The Act forbade private resource ownership, preventing non-residents from controlling it. The Act was based on the claim that the state, not indigenous communities, should own the resource. Open access replaced effective common property management, leading to resource degradation and requiring formal regulation.

- A long-running dispute in the U.S. tests the limits of private property rights. The legal doctrine of "regulatory taking" seeks to define zoning and environmental rules that diminish the value of property as "takings", requiring compensation under the Fifth Amendment to the U.S. constitution. The doctrine's objective is to weaken governments' police powers (e.g. environmental regulation), strengthening the rights of private property owners.[1] Periodic legal disputes in coastal states test landowners' ability to impede access to beaches, or to insist on public investment (e.g. sea walls) that maintains the value of private property.

Private property diminishes or eliminates some common property or open access externalities. Grazing an additional cow on a field creates benefits for the cow's owner. If the cow competes with other animals for fodder, it creates a negative externality for other users, much as an additional driver contributes to road congestion. The overgrazing also damages the field, lowering its long run productivity, thus lowering both short and long run community welfare. This outcome is known as the "tragedy of the commons". Private owners internalize the congestion created by the additional cow, and are therefore less likely to overgraze the field. If private property solves the tragedy of the commons, it improves resource management.

An emerging body of research shows that many societies have successfully managed common property natural resources, avoiding the tragedy of the commons. Common property management requires widespread agreement on the rules of use, and mechanisms for monitoring and enforcement of the rules. Stable conditions and homogenous users increase the success of common property management. A rapid change that increases the demand for the resource, such as migration or opening to trade, can undermine common property management. Both private and common property require monitoring and enforcement.

Events during the final decade of the 20th century illustrate that private property does not guarantee efficient management. When the Soviet Union collapsed in 1991, reformers and their western advisors (primarily economists) debated the right pace of privatization of state owned property. Those supporting rapid privatization hoped that it would lead to the efficient use of natural and man-made capital, and feared that a slower pace would only perpetuate the inefficiencies and make it possible to reverse the reforms. Privatization occurred rapidly, but instead of leading it efficiency it created a class of oligarchs and an entrenched system of corruption.

[1] Supreme Court rulings have largely reaffirmed these police powers, undermining the Doctrine of Regulatory Takings. Bilateral Investment Treaties (Chapter 11.2) requiring compensation for regulation that is "tantamount to expropriation" strengthens the Doctrine in the sphere of international, rather than US domestic law.

Following political upheaval, local elites might capture resources, replacing common property with private property, rendering the traditional monitoring and enforcement mechanisms irrelevant. In a politically unstable environment, the new owners recognize that the next political upheaval might replace them with another group of elites, making their property rights insecure. The combination of current absolute but insecure property rights is particularly likely to lead to overuse of the resource. The current owners can use the resource to enrich themselves; the risk of losing control of the resource causes them to attach little weight to its future uses.

In summary, actual property rights tend to exist on a continuum that includes the three main types as special cases. Regulation is particularly necessary under open access, which is especially vulnerable to the tragedy of the commons. Common property management in small communities has (often) avoided this tragedy. As communities integrate into wider markets, the management practices frequently break down, requiring different kinds of regulation. Private property solves some externality problems but creates others, and also typically requires regulation. For example, converting the village commons to a privately owned farm gives the farmer the incentive to manage the land efficiently (e.g. to avoid erosion), but not to correct off-farm externalities (e.g. pollution run-off). Arguably, the problem here lies not with private ownership of the land, but with the lack of property rights to the waterways that absorb the pollution. With this view, the policy prescription is to create property rights for the waterways. Creating those additional rights may be too expensive or politically or ethically unacceptable, in which case the policy prescription is to regulate pollution.

12.2 The Coase Theorem

Objectives and skills

- State and explain the Coase Theorem.

Transactions costs include the costs of reaching and enforcing an agreement. The Coase Theorem states that if transactions costs are negligible, and property rights are well-defined, agents can reach the efficient outcome regardless of the *distribution* (or allocation) of property rights. Rational agents will not leave money on the table. Under the conditions of the theorem, there is no need for a regulator, because private agents reach an efficient outcome by bargaining. The government's only role is to insure that agents honor their contracts, and possibly to make transfers in order to promote fairness; efficiency does not require those transfers. In many situations where we observe regulation instead of a bargained outcome, transactions costs are large, making it impractical for agents, bargaining amongst themselves, to achieve an efficient outcome. In other situations, the property rights are not well-defined, leaving agents uncertain about the payoffs of reaching a bargain, making an agreement harder to reach.

To illustrate the Coase Theorem, suppose that the total profit of a fishery depends on the number of boats operating there. With one boat, the profit is 1, with two boats, the aggregate profit is 4, and with three boats, the aggregate profit is 3. Initially, three boats, each with a separate owner, operate in the fishery. The surplus obtained from inducing one boat to exit equals $4 - 3 = 1$. A person with exclusive property rights to the fishery would insist that the other two fishers leave the sector, and would then buy an additional boat, reaching the efficient outcome. If all three fishers have some property rights, and if the transactions costs are small, then they can reach an agreement in which one of them sells her right to fish to the other two and leaves the sector. There are many types of bargains that they might strike, leading to different splits of the surplus.

For example, if the three fishers are equally productive, they can create a lottery that determines who leaves the sector.[2] The person who leaves receives a payment of $1 + x$, their initial profit plus the compensation x for leaving. The two remaining fishers split the higher profit, each receiving 2, and share the cost of buying out the departing fisher, for a net benefit of $2 - \frac{1+x}{2}$. The expected payoff to an agent participating in a fair lottery (where the chance of leaving the sector is 1/3) is $\frac{1}{3}(1 + x) + \frac{2}{3}\left(2 - \frac{1+x}{2}\right) = \frac{4}{3}$, which is greater than their payoff under the status quo (1). If $x > \frac{1}{3}$, the person who leaves is a winner, and if the inequality is reversed, this person is a loser; setting $x = \frac{1}{3}$ insures that there are no losers. Other procedures (e.g. arm wrestling) could also be used to determine who leaves.

The Coase Theorem does not predict how the efficient outcome is obtained (a lottery or arm wrestling) or the compensation (x in the lottery example). It merely says that if transactions costs are small and property rights well defined, rational agents will bargain and achieve an efficient outcome. Dif-

[2] If the fishers are not equally productive, then an efficient procedure must choose the two most productive to remain in the fishery.

ferences in bargaining power, perhaps due to differences in outside options (a fisher's best alternative to remaining in the sector) or levels of bargaining skill lead to different *distributions* but not different aggregate bargaining surplus.

12.3 Regulation of fisheries

Objectives and skills

- Know the basics of the recent history of fishery regulation.
- Understand limitations of current regulations, and the effects of assigning property rights.

Prior to the 20th century, the doctrine of "freedom of the seas" limited nations' sovereignty to three miles from their coastline, permitting other nations to operate outside that area. In the 20th century, countries began to claim sovereignty over larger areas, often to protect their fisheries. The United Nations Convention on the Law of the Sea, concluded in 1982, replaced earlier agreements, giving nations an "exclusive economic zone" (EEZ), e.g. to harvest fish or extract oil, 200 miles beyond their coastline.

The U.S. passed the Manguson-Stevens Fishery Conservation and Management Act in 1976 and amended it in 1996 and 2006, to manage fisheries within its EEZ. The goals included: conserving fishery resources, enforcing international fishing agreements, developing under-used fisheries, protecting fish habitat, and limiting "bycatch" (fish caught unintentionally, while in pursuit of other types of fish). The law established Regional Fishery Management Councils, charged with developing Fishery Management Plans (FMPs). These FMPs identify overfished stocks and propose plans to restore and protect the stocks. The law requires management practices to be based on science. For each fishery, a scientific panel determines the "acceptable biological catch", and the managers then set an "annual catch limit" (ACL), not to exceed the acceptable biological catch. The Fishery Management Councils can enforce the ACL by: limiting access to specific boats or operators; restricting fishing to certain times of the year or certain locations; regulating fishing gear; and requiring on-board observers to insure that boats obey regulations. A 2009 National Marine Fisheries report states that of the 192 stocks being monitored, the percent with excessive harvest fell from

38% to 20% over the decade, and the percent with over-fished stocks fell from 48% to 22%.

Most of the world's important fisheries are concentrated in countries' EEZs and are regulated in some fashion. However, political constraints, scientific limitations, and human frailty often results in over-fished stocks. Identifying the actual stock is a difficult measurement problem. Random stock changes, due to unforeseen ocean conditions and changes in stocks of predators and prey make it difficult to determine safe stock levels. Managers may decide to ignore scientific evidence if they distrust it or are swayed by lobbying from fishers or processors. "Regulatory capture" occurs when regulators substitute society's goals with the narrower interests of the groups that they are charged with regulating.

Box 12.1 The U.S. Northwest Atlantic scallop fishery: a success story. Scallops are caught by dragging dredges along the seabed. In 1994 the U.S. government closed 3 large banks, causing scallopers to move to and subsequently exhaust other areas of the fishery. Fisherman hired University of Massachusetts biologists to conduct a stock survey, which concluded that the population in the closed areas had rebounded. The industry began to commit a fraction of its profits to conduct surveys. Using this data, the National Marine Fishery Services (NMFS) closes overfished seabeds long enough to allow the population to recover (about 3 - 5 years), a system resembling field rotation in agriculture. The success of the program relies on good data and cooperation between fishers and NMFS. Fishers have a common interest in preserving the stock, and they trust the data because they supply it.

The mismatch between "targets" (or "margins") and "instruments" (or policies) complicates fishery regulation (Chapter 9). A target is anything that a regulator would, in an ideal world, like to control. The payoff from a fishery involves many considerations in addition to the annual catch limit, including specifics about nets, engine size, and other boat equipment. In principle, the Regional Fishery Management Councils have the authority to regulate most of these features, but regulation of every important business decision is not practical. Most regulation involves fairly simple policies, e.g. restricting the length of the season or limiting the number of boats.

When the regulator restricts one margin (e.g. by limiting the season length), fishers respond on another margin, (e.g. by changing the gear they use). Regulators cannot control all margins, so regulation is second-best. Fishers acting in their self-interest inflict externalities on others, leading to inefficient outcomes. This inefficiency often manifests as over-capitalization: too many boats, or too much gear, chasing too few fish. Over-capitalization reduces industry profits, making a given annual catch limit more expensive to harvest. Estimates of over-capacity range from 30% to over 200%, but there is a consensus that it is severe.

12.3.1 A model of over-capitalization

We illustrate over-capitalization and the loss in industry profits using a model with four targets: annual allowable catch, A; the number of boats, N; the number of days of the season, D, and the amount of "effort" per boat per day, E. Effort is an amalgam of all of the decisions the fisher makes (boat size, crew, gear characteristics). Each unit of effort costs w per day, and E units of effort enable a boat to catch $E^{0.5}$ fish per day.

In the first best setting, the regulator chooses the value of all four variables. To examine the forces at work in the real world, where regulators are unable to control every margin, we consider a second best setting where the regulator chooses the annual catch, A, and enforces that choice by selecting the length of the season, D. Fishers choose effort, E, to maximize their profits. We first consider the case where N is fixed, and then the case where N is endogenous.

Exogenous N

Individual fishers, each with a boat, choose effort in order to maximize profit (revenue minus cost). If the price of a unit of fish is p, the fisher chooses E to maximize profits per day:

$$\max_E \left(pE^{0.5} - wE\right) \Rightarrow \frac{d\left(pE^{0.5} - wE\right)}{dE} \overset{set}{=} 0 \Rightarrow E = \left(0.5\frac{p}{w}\right)^2. \qquad (12.1)$$

Given that the N fishers choose this level of effort, the manager who wants to set annual catch equal to A must choose the number of days of the season, D, so that the number of fishers (N) times the catch per day per fisher ($E^{0.5}$) times the number of days equals A:

$$NE^{0.5}D = N\left(\left(0.5\frac{p}{w}\right)^2\right)^{0.5} D \overset{set}{=} A \Rightarrow D = 2\frac{Aw}{Np}.$$

The maximum feasible season length (e.g., due to weather conditions) is S. We denote $\delta = \frac{D}{S}$, the fraction of the potential season that fishing is allowed, and we assume that $\frac{D}{S} < 1$ i.e.,

$$\delta = \frac{2Aw}{NSp} < 1. \tag{12.2}$$

This inequality implies that, given fishers' individually optimal choice of effort per day, the annual allowance, A, constrains the length of the season. If inequality 12.2 were reversed, then fishers could work the entire season without exceeding the ceiling, making regulation unnecessary.

In this second-best setting, total industry profit equals

$$\text{profit}^{\text{second best}} = DN\left(pE^{0.5} - wE\right) =$$
$$2\frac{A}{N}\frac{w}{p}N\left(p\left(\left(0.5\frac{p}{w}\right)^2\right)^{0.5} - w\left(0.5\frac{p}{w}\right)^2\right) = \frac{Ap}{2}$$

In the first best setting, the regulator chooses both effort and the number of days. The fixed harvest fixes revenue, so profit maximization requires minimizing the cost of harvesting A. Cost minimization requires setting $D = S$, the maximum feasible season. The increase in D relative to the second best setting, requires a reduction in effort per day. The first best level of effort is $E = \left(\frac{A}{NS}\right)^2$ and total profit in the industry is

$$\text{profit}^{\text{first best}} = DN\left(pE^{0.5} - wE\right) =$$
$$SN\left(p\left(\frac{A}{NS}\right) - w\left(\frac{A}{NS}\right)^2\right) = pA - \frac{A^2w}{NS}.$$

The percent increase in profit in moving from the second best to the first best regulation is

$$\frac{\text{profit}^{\text{first best}} - \text{profit}^{\text{second best}}}{\text{profit}^{\text{second best}}}100 = \frac{\frac{1}{2}Ap\left(1 - \frac{2Aw}{NSp}\right)}{\frac{Ap}{2}}100 = (1-\delta)\,100 > 0.$$

Fishers act in their self-interest in choosing effort, but they inflict a negative externality on the industry: as effort increases, catch per day also increases, requiring that the regulator shorten the season in order to maintain the annual catch limit. This self-interested behavior leads to $\frac{1}{\delta^2}$ times the first best level of effort. This static model does not distinguish between a durable input, such as a boat, and a variable input, such as effort. In this context, we take "over-capitalization" to mean that there are too many productive assets in the sector. Over-capitalization decreases industry profits. If second best regulation leads to a season only 20% as long as the first best level, then industry profits are only 20% of the first best level.

This model illustrates an empirically important phenomenon: the race to catch fish (and resulting over-capitalization) leads to shorter seasons and lower industry profit in many fisheries. In the early 20th century, the North Pacific halibut fishery operated throughout the year, leading to excessive harvests. In 1930 the U.S. and Canada agreed to manage the fishery using an annual quota. The initial management success increased profits, which lead to a larger fleet. Managers responded by reducing the season length to two months in the 1950s and to less than a week in the 1970s.

Endogenous N

If the number of boats, N is endogenous, it responds to conditions in the fishery, and thus is another "target", or "margin". Higher profits encourage entry. Suppose that initially the fishery is totally unregulated, leading to depleted stocks, making it expensive to catch fish. The high costs lead to low profits. In equilibrium, boats earn their opportunity cost; no boats want to enter or leave the sector. Now suppose that a restriction on annual catch succeeds in rebuilding the stock, making it cheaper to catch fish, and thus increasing profits. In the absence of entry restrictions, more boats may enter, creating another type of over-capitalization.

Suppose that each boat costs $\$F$/year to own; F is the opportunity cost of the money tied up in owning a boat for a year. Under second best regulation, where the length of the season adjusts in order to maintain the catch limit, we saw above that industry profit is $\frac{Ap}{2}$. With N boats, the profit per boat per year is $\frac{Ap}{2N}$. Under free entry, boats would enter the industry until the annual opportunity cost of being in the sector equals the profit of fishing there, which requires $\frac{Ap}{2N} = F$, or $N = \frac{Ap}{2F}$. (This calculation ignores the fact that the number of boats must be an integer.) For smaller values of N additional boats have an incentive to enter the fishery (to obtain positive net profits), and for larger values of N existing boats have an incentive to leave the fishery (to avoid losses).

Consider the case where the number of boats is in a second best equilibrium ($N = \frac{Ap}{2F}$). A change suddenly enables the regulator to control effort, determining the gear each boat uses. We showed above that this change, at the initial level of N, increases industry profit by $(1-\delta)\,100\%$, increasing the profit per boat by $\frac{(1-\delta)100}{N}\% = \frac{(1-\delta)2F100}{Ap}\%$ (using the expression for the equilibrium level of N). This increase in profit encourages additional boats to enter.

The second best equilibrium number of boats ($N = \frac{Ap}{2F}$) is already excessive from the standpoint of society, but at least the regulator does not have to discourage more boats from entering: they have no desire to do so. However, successfully reducing effort per boat creates an incentive for new boats to enter. Getting one margin right (reducing effort per boat) causes another margin (the number of boats) to move further from the social optimum. This kind of problem is endemic to a second best setting: fixing one problem can make another problem worse. In principle, the regulator can determine the optimal number of boats and the optimal effort per boat, and then either decree that fishers accept these levels, or impose taxes (e.g. a tax per unit of effort and an entry tax) that "induce" these levels. That level of regulation is seldom practical.

Box 12.2 The relation between N and effort. In the model here, fishers' equilibrium choice of effort does not depend on the number of boats (equation 12.1). That independence simplifies the calculations, but it is a consequence of functional assumptions. More generally, equilibrium effort depends on the number of boats. Regulating one aspect of the industry affects other fishing decisions. Economic agents who are constrained in one dimension (e.g. not being able to increase the number of boats) respond by making changes in other dimensions (e.g. increasing the speed of boats). Over-capacity can show up in many ways: as too many boats, too much gear, or boats that are too big or too fast.

12.3.2 Property rights and regulation

When individuals fishers pursue their self interest, their aggregate decisions often lead to over-fishing, resulting in low stocks and low industry rent. Because individual fishers do not own the stocks, they have little incentive to manage them for future users. Even if the over-fishing problem can be solved by setting and enforcing an annual catch limit, history shows that other inefficiencies remain, often leading to over-capitalization and low profits.

Creating property rights is the obvious alternative to regulation such as limiting the fishing season, restricting the number of boats, and regulating fishing gear. Problems in fisheries arise largely from the lack of property rights, so creating these rights is the most direct remedy; recall the Principle of Targeting from Chapter 9. Individual Quotas (IQs), or Individually Transferable Quotas (ITQs) are the most significant form of property rights in fisheries. These measures set an annual Total Allowable Catch (TAC) for a species and give (or sell) individuals shares of that quota. An individual with share s is entitled to harvest s times that year's TAC of that species. Under ITQs, owners can sell or lease ("transfer") their shares.

Property rights-based regulation has been shown to decrease costs, increase revenue, and protect fish stocks. However, less than 2% of the world's fisheries currently use property rights based regulation.

Reducing industry costs For a given TAC (or ACL), property rights cause fishers to internalize the externalities that lead to overcapitalization. Property rights reduce costs, increasing profits. Using the model above, a fisher with the share s of a TAC A for a fishery with price p obtains the fixed revenue sAp. With fixed revenue, the fisher maximizes profits by choosing effort and days fished to minimize the cost of catching sA. This fisher's cost minimization problem is the same as that of the regulator who chooses both effort and season length. Thus, the fisher with property rights chooses the socially optimal amount of effort and days fished.

Transferable quotas (ITQs instead of IQs) provide two additional avenues for cost saving. First, some of the fishers may be more efficient. The average fisher may catch $E^{0.5}$ fish per day, but half may catch 10% more and half 10% less with the same level of effort. Regulators are unlikely to know fishers' relative efficiency, or be able to act on that information even if they have it. Fishers probably have a better idea of their relative efficiency. Shares in the quota are worth more to the efficient fishers than to the inefficient fishers, so there are opportunities for the former to buy the latter's' licenses. If that occurs, aggregate costs in the fishery falls by 10% in this example. Second, even if all fishers are equally productive, there may be too many boats in the sector: industry profits would be higher if some boats could be persuaded to leave, as in the example in Chapter 12.2. If transactions costs are sufficiently low, and fishers manage to solve the Coasian bargain, some boats sell their quotas and leave the sector. Profits under the smaller fleet are higher.

The Mid-Atlantic surf clam fishery switched from restricted entry regulation to ITQs in 1990. Prior to the switch, there were 128 vessels in the fishery. Estimates at the time claimed that the optimal number of boats for the industry was 21 – 25, and that rationalization of effort could reduce costs by 45%. By 1994, the fleet size had fallen to 50 vessels and costs had fallen by 30%.

Increasing revenue For a fixed TAC, IQs and ITQs can increase revenue, in addition to reducing harvest costs. Partial regulation leads to too much gear and/or too many boats chasing a given number of fish, requiring regulators to reduce the fishing season, sometimes to a period of a week or less. Over-capitalization increases fishing costs, but it also makes the annual harvest available during a short period of time, instead of being spread out during the season.

If the market for fresh fish is inelastic, a sudden increase in harvest leads to a much lower equilibrium price. There may also be capacity constraints in transporting fresh fish to market. Fish processors may be able to wield market power if fishers have to unload large catches during short periods of time. For all of these reasons, the ex vessel price that fishers receive may be lower when the annual catch is landed during a short interval. Moving from regulation to IQs or ITQs eliminates or at least reduces the race to catch fish, causing harvest to be spread out over the season, and increasing the average price of landings.

Before the British Columbia halibut fishery switched to ITQs in 1993, the fishing season lasted about five days, and most of the catch went to the frozen fish market. With the more spread-out and thus steadier supply of fish caused by the move to ITQs, wholesalers found it profitable to develop marketing networks for transporting fresh fish. Prior to the ITQs, the price of fresh fish fell rapidly if more than 100,000 pounds a week became available, but with the development of the new marketing networks, the market could absorb 800,000 pounds before prices dropped.

A 2003 study estimates the potential gains from introducing ITQs in the Gulf of Mexico reef fish fishery, accounting for both cost reductions and revenue increases. ITQs had the potential to increase revenue by almost 50% and to reduce costs by 75%. The study also estimates that under ITQs the equilibrium fleet size contains 29 – 70 vessels, compared to the actual level of 387 vessels.

Protecting fish stocks Standard regulation and IQs/ITQs both rely on an annual limit (TAC) but they use different ways of enforcing that limit. The mismatch between targets and instruments under standard regulation, and the resulting inefficiencies, lead to high costs, low revenue, and financial difficulties for fishers. Fishers have an incentive to pressure regulators to increase the annual catch to provide short term financial relief. If this pressure overrides scientific advice, it imperils fish stocks, making the longer term problem worse. The creation of property rights potentially eases these stresses. Fishers with property rights to the catch have an incentive to protect the stock. Property rights-based regulation potentially changes the political dynamics, helping to protect fish stocks.

The difficulty of measuring stocks makes it hard to know which fisheries are at risk. A common definition calls a fishery "collapsed" in a particular year if the harvest in that year is less than 10% of the maximum previous harvest. By this definition, 27% of the fisheries were collapsed in 2003. The data shows that fisheries with IQ/ITQs were less likely to be collapsed, but the "selection problem" makes it difficult to tell whether this negative correlation between ITQ status and fishery collapse is spurious, or whether the property rights-based mechanism really protects the fishery's health. The problem is that the econometrician does not observe the selection process that determines whether a fishery is managed by IQ/ITQs or by some other means. Suppose that political considerations make it easy to convert some fisheries to property rights-based management, and difficult to convert others; suppose also that the "politically easy" fisheries happen to be less prone to collapse. These two circumstances tend to create a negative correlation between property rights-based management and collapse status, independently of whether there is a causal relation between the two. Statistical methods based on "matching" can alleviate this measurement problem. The idea is that we would like to compare collapse status between pairs of fisheries that are alike, except for their ITQ status.

A 2008 study based on 50 years of data and over 11,000 fisheries, taking into account the selection problem, estimates that ITQs reduce the probability of collapse in a year by about 50%. It also estimates that had there been a general movement to ITQs in 1970, the percent of fisheries collapsed in 2003 would have been about 9% instead of the observed 27%. In contrast, a 2009 study based on 18 countries (with over 100 fish stocks and 249 species) found that fish stocks continued to decline in eight of the 20 stocks regulated using ITQs. ITQs are not a panacea.

Remaining issues The property rights created by IQs and ITQs can improve the financial and ecological health of a fishery, but they leave many problems unsolved. They create particular property rights to a particular species, but may leave important externalities. Increases in production costs during the season due to the intra-seasonal decrease in stock, can still create a race to catch fish, leading to overcapitalization. The IQs and ITQs do not protect other species. They provide no incentive to reduce by-catch, the unintentional harvest of fish. Other regulations, or the creation of additional property rights can reduce by-catch, but these may be costly to implement and may have unintended consequences. A fairly new mechanism, Territorial Use Rights Fisheries (TURFs) attempt to alleviate the cross-species problem by giving groups of fishers (e.g. coops) exclusive rights to an area, instead of to a species. TURFs are less effective if important species move in and out of the territorial area.

ITQs likely lead to industry concentration, reducing the number of boats and fishing jobs. Objecting to ITQs because they harm fishing communities is unconvincing for two reasons. First, overcapitalization is a major problem in the fishing sector. There is no way to remedy this problem without decreasing the size of the sector, which includes reducing employment. Second, even though reducing the size of sector can create real and sometimes long-lasting hardship, the recommendation to keep an industry inefficiently large, in order to support local employment, is not persuasive in general. The "local employment" argument is routinely used as a rationale for supporting shrinking industries, including: steel production in the U.S. during the 1980s and 90s; forestry in the U.S. during the 1990s and 2000s; and agriculture (e.g. in the European Common Agricultural Policy). The Principle of Targeting tells us that even though local unemployment may be a significant social problem, maintaining an inefficiently large sector is unlikely to be the right policy prescription. For natural resource-based industries, this employment argument is particularly unpersuasive. Supporting employment in the sector aggravates the decline in resource stocks. Unless something is done to protect the resource upon which the sector relies, employment in the sector will certainly fall.

A second possible reason for being concerned about the increased concentration (caused by ITQs) is that it might make it easier for fishers to form a cartel and exercise market power. However, even if a small number of fishers account for a large fraction of catch in a particular fishery, they face significant competition from other fisheries and other food products. Their ability to exercise market power is likely small. ITQs might also be criticized because they transfer resource rents to individual fishers, instead of society at large. However, this is an argument against the way in which ITQs are distributed (as a gift rather than by auctioning), not against ITQs as a means of protecting a resource stock.

12.4 Subsidies to fisheries

Policy failure harms natural resources by permitting and sometimes encouraging over-harvest, reducing stocks. This loss in natural capital threatens future harvests, and the short run economic effects include over-capitalized and financially stressed fisheries, and falling supplies of high-value catch. The policy remedy requires reducing catch in order to allow stocks to recover, and encouraging rationalization and consolidation of the industry. Reducing catch and encouraging fishers to leave the industry are politically unpopular.

Subsidizing the industry is politically easier, but ultimately counterproductive. Subsidies disguise the economic costs, enabling fishers to remain in an unprofitable activity, worsening both the problem of over-harvest and over-capitalization. Both domestic and international agencies have documented the link between subsidies and over-fishing and over-capitalization. A 2013 European Parliament study estimates annual global subsidies to the fishing sector of \$35 billion (2009 dollars). Over half of those subsidies generate increased capacity; 22% come in the form of fuel subsidies, 20% subsidize fishery management, and 10% subsidize ports and harbors (Table 12.1). Developed countries are responsible for most of the subsidies; 43% originate in Asia, chiefly Japan and China. Between 1996 – 2004, the U.S. fishing industry received over \$6.4 billion in subsidies.

type of subsidy	fuel: 22	management: 20	ports and harbors:10	capacity increasing: 57
source of subsidy	developed countries: 65	Asia: 43 Japan: 20 China: 20	Europe: 25	North America: 16

Table 12.1 Types and geographical sources of fishing subsidies, as a percent of total. (Sumaila et al. 2013)

Subsidies transfer income from one group to another, here from tax payers to fishers, and indirectly to consumers via lower prices; subsidies also create distortions by attracting mobile factors of production into a sector. In the static setting with a single distortionary tax (or subsidy), Chapter 10.3 notes that the economic cost of the policy (the deadweight loss) is typically much *smaller* than the transfer. ("Triangles are small relative to rectangles.") The example in Chapter 9.5 shows, however, that mutually reinforcing distortions can lead to much higher social losses, exceeding the magnitude of the transfer. That example shows that a subsidized sector might contribute negative value added to society: the social value of the mobile inputs used in the sector exceeds the social value of production.

A 2009 study commissioned by the World Bank estimates the annual global economic cost of fishery subsidies at $50 billion, which is larger than the fiscal cost of the subsidies. This economic cost includes the costs of over-capitalization and the cost due to reduced stock. In some fisheries, the value of harvest is less than the true cost of harvest, as in the example in Chapter 9.5. Here, the industry operates at a "social loss", disguised by government subsidies.[3] The same study estimates that half the current number of vessels could achieve current catch. Subsidies account for approximately 20% of fishing revenue.

Box 12.3 Subsidies in other sectors. Fisheries are not alone in receiving politically motivated but economically unproductive subsidies; Chapter 11.1 discusses fossil fuel subsidies. Agriculture in many developed countries also receives large subsidies, often justified as helping struggling farmers. However, endogenous changes induced by the subsidies often undo whatever short term financial help the subsidies provide. Agricultural subsidies are "capitalized" in the price of land: the expectation that the subsidies will continue into the future make people willing to pay more for land, raising its equilibrium price. Current land owners who sell their land, not entering farmers who must buy the land, capture these increased rents. Subsidies, and in particular the belief that they will continue into the future, increase young farmers' debt (via the increase in land prices), making them more vulnerable to future financial difficulties and more dependent on the subsidies.

[3] This situation is reminiscent of many Russian and east European industries after the collapse of the Soviet Union. These industries were not economically viable, and had been kept alive by government subsidies; closing them down increased gross national product.

12.5 Summary

Private property, common property, and open access are the three leading modes of property rights. Property rights to most resources lie somewhere on the continuum that includes these three modes. *De facto* property rights sometimes differ from *de jure* rights. Common property can lead to overuse of the resource, a result known as the tragedy of the commons. Many societies developed mechanisms that efficiently manage common property resources; increased trade, migration, and population growth sometimes erode these mechanisms. If transactions costs are small and property rights well defined, agents can (plausibly) reach an efficient outcome through bargaining. The Coase theorem states that in this case, the efficiency of the outcome does not depend on agents' bargaining power, or more generally, on the assignment of property rights. This result implies that regulation is unnecessary when transactions costs are low and property rights are secure.

Most economically important fisheries fall within nations' Exclusive Economic Zone, and are regulated. The impracticality of regulating every facet of fishing leaves regulators in a second best setting. Many fisheries set annual quotas, enforced using a variety of policies, notably early season closures. The race amongst fishers to catch fish leads to over-capitalization, which results in high costs and, because much of the harvest is landed during a short period of time, low revenue. Property rights-based regulation, primarily ITQs, can lead to consolidation and rationalization of fisheries, lowering costs, increasing revenue, and via a political dynamic, increasing the prospect for fishery health (adequate stocks). However, only a small percent of fisheries operate under ITQs. Subsidies, especially from developed nations, provide short run benefits to fisheries, but often exacerbate the causes of low profits and over-fishing.

12.6 Terms, study questions and exercises

Terms and concepts

Open access, common property, *de jure, de facto*, congestion, tragedy of the commons, Doctrine of Regulatory Taking, transactions costs, Coase Theorem, fair lottery, outside option, Law of the Sea, exclusive economic zone, Manguson-Stevens Fishery Act, by-catch, Regional Fishery Management Councils, Fishery Management Plans, acceptable biological catch, annual catch limit, regulatory capture, overcapitalization, Individual (Transferable) Quotas or I(T)Qs, Territorial Use Rights Fisheries (TURFs).

Study questions

1. Recently there have been a spate of disputes on airlines concerning whether a person is entitled to recline their seat. These disputes are essentially about who has the property rights to the several inches of space between a seat and the one in front of it. Some airlines have (either implicit or explicit) rules that assign property rights: a person is entitled to recline her seat, except during meals (and of course during landing and takeoff). Recently these rules seem to have become vaguer, or less well understood. Discuss the increasing occurrence of these disputes among passengers, in light of the Coase Theorem.

Exercises

1. This exercise takes the reader through one of the classic examples of the Coase Theorem. A factory that emits e units of pollution obtains the total benefit from emissions, $10e - \frac{1}{2}e^2$. A (very old fashioned) downstream laundry dries clothes outside. The pollution makes it more expensive for the laundry to return clean clothes to its customers, and therefore increases the laundry's costs by $2e^2$. (a) Find the socially optimal level of pollution, i.e. the level that maximizes benefits minus costs. (b) Find the emissions tax that supports this level of pollution as a competitive equilibrium. (c) Assume that the factory has the right to pollute as much as it wants. The laundry and the factory are able to costlessly bargain to reduce pollution. Who pays and who gets paid in the bargaining outcome? (d) Justify the Coasian conclusion, namely that the outcome of an efficient bargain leads to the socially optimal level of pollution. One can establish this claim using the following proof by contradiction: Suppose, contrary to the claim, that they reach a bargain that entails an amount of pollution different than the socially optimal level. (To avoid repetition, consider only the case where this amount is greater than the socially optimal level.) Show that at such a level, the laundry's willingness to pay for a marginal reduction in pollution is strictly more than the firm would have to receive in order for it to be willing to reduce pollution by a marginal amount. Explain

why this conclusion implies that if the agreement by the factory and the laundry leads to more pollution than is socially optimal, then the factory and the laundry have left money on the table. They should resume bargaining and capture some of this waste.

2. Chapter 13.3.1 shows the representative firm's optimization problem and the equilibrium level of E. Parts (a) and (b) of this question primarily involve copying, together with filling in a couple of details. Part (c) requires following a series of instructions. Part (d) requires summarizing the interpretation given in the text. (a) Copy the optimization problem, write down the first order condition, and solve it to obtain the value of E shown in equation 13.1. (b) Write down the manager's constraint and solve it to find the value of D. (Essentially, just copy the equation shown below the numbered equation 13.1 and solve it to obtain the expression for D. (c) Obtain the formula for effort under "full regulation" (where the regulator is able to choose both D and E), given in Chapter 12.3.1. Because harvest cannot exceed A, the total revenue is fixed at pA. Therefore, the regulator's objective is to minimize costs, subject to the constraints that the limit is caught ($NE^{0.5}D = A$) and that D not exceed the maximum number of days, S, during which (e.g. due to weather conditions) it is feasible to fish ($D \leq S$):

$$\min_{E,D} (wED) \text{ subject to } NE^{0.5}D = A \text{ and } D \leq S$$

Proceed as follows: (i) Use the constraint involving A to solve for E. (ii) Substitute this value of E into the minimand (the thing being minimized, here, costs). (iii) Note that the resulting minimand is decreasing in D. Conclude that the value of D that minimizes costs is therefore the maximum feasible value, S. (iv) Using $D = S$ from part (iii) in the expression you obtained from part (i), write the level of effort under "full regulation" as a function of the model parameters. (d) Explain why the values of E and D are different under "partial regulation" (where fishers choose E and the manager chooses D), versus "full regulation", where the manager chooses both variables.

3. Using the model in Chapter 12.3.1, and the formulae given there, show that under partial regulation fishers choose $\frac{1}{\delta^2}$ times as much effort per day as the regulator chooses under full regulation.

4. Use the model in Chapter 12.3.1. (a) What is the economic meaning of the inequality $\frac{2Aw}{Sp} < 1$. (Hint, look at the formula for δ.) (b) The text gives the formula for industry profits, (profit$^{\text{second best}}$) (ignoring the cost of boats), when fishers choose effort. Does this level of profit depend explicitly on N? (c) Suppose that the inequality $\frac{2Aw}{p} < S$ holds, and suppose also that each boat costs \$$F$ per year. What is the socially optimal number of boats when fishers choose the level of effort? (d) Now suppose that the regulator chooses the the amount of effort. Ignoring the cost of a boat, the text gives the expression for industry profits in this case, profit$^{\text{first best}}$. If each boat costs \$$F$ per year and there are N boats, what is the industry profit, net of the cost of the boats? (Just write down the payoff.) (e) What is the first order condition (with respect to the choice of N) for maximizing industry profit in this case? (f) What is the necessary and sufficient condition for the optimal number of boats to exceed 1? (g) Why does optimal number of boats (typically) differ in the two cases (where fishers choose effort, and where the regulator chooses effort)?

Sources

Dietz, Ostrom and Stern (2003) and Ostrom (1990 and 2007) discuss conditions under which societies successfully manage common property without formal regulation.

Hardin (1968) introduced the term "tragedy of the commons".

Gordon (1954) provided the first well known analysis of fisheries as common property resources.

Kaffine (2009) documents California surfers' efforts to exercise de facto property rights on some beaches.

The Millennium Ecosystem Assessment (United Nations, 2005), Sumaila et al. (2011) and Dyck and Sumaila (2010) provide overviews of the state of fisheries.

Johnson and Libecap (1982) describe the replacement of common property with open access in the Alaska salmon fishery.

Fenske (2012) documents the case of property rights for rubber trees among the Nigerian Igbo.

The 2009 National Marine Fisheries Service report to Congress summarizes the change in regulated U.S. fisheries.

Wittenberg (2014) provides the information for Box 12.1.

Smith (2012) surveys the problems of regulating fisheries when there more "targets" (aka "margins") than policy variables.

Homans and Wilen (1997) show how individually rational effort decisions, and the resulting overcapitalization, lead to reductions in the length of a fishing season.

Homans and Wilen (2005) document the effect of ITQs on revenue, and provide the example of the British Columbia halibut fishery.

Weniger and Waters (2003) estimate potential revenue gains, cost reductions, and fleet consolidation due to using ITQs in the Gulf of Mexico reef fish fishery.

Weninger (1999) provides the information in Chapter 12.3.2 on the Mid-Atlantic surf clam.

Abbott and Wilen (2009 and 2011) examine fishers' response to quotas on bycatch.

Costello, Gaines and Lynham (2008) estimate the effect of ITQs on fishery collapse, reported in Chapter 12.3.2.

Deacon, Parker, and Costello (2013) study a situation where fishers had the option of obtaining property rights by joining a coop.

The 2012 Symposium "Rights-based Fisheries Management", edited by Costello and with papers by Aranson, Deacon, and Uchida and Wilen, reviews the literature on rights-based management.

Sharp and Sumaila (2009) quantify U.S. fishery subsidies.

Sumaila et al. (2013) quantify global fishery subsidies.

Aronson, Kelleher and Willman (2009) estimate the economic cost of global fishery subsidies.

Chapter 13

Renewable resources: tools

Objectives

- Introduce the building blocks of renewable resource models.

Skills

- Understand and be able to work with a growth function.
- Understand the meaning of a "harvest rule".
- Know the meaning of a steady state, and understand the relation between a growth equation and a steady state.
- Understand the meaning of stability, and be able to test for it in a continuous time model.
- Understand the meaning of "maximum sustainable yield" and be able to identify it for simple growth functions.

A few basic tools make it possible analyze a range of renewable resources. In subsequent chapters we use these tools to study the open access fishery and the sole owner fishery. We then use the tools to study water economics, where we also note their broad applicability. Resource models involve one or more "stock variable(s)" that (potentially) change over time. In the nonrenewable resource setting, the stock variable equals the amount of the resource remaining in the mine; there, any extraction decreases the stock. In the renewable resource setting, growth might offset extraction, causing the stock to either decrease or increase over time.

In water economics, we might measure the stock using the level of groundwater or the amount of water in a reservoir. Consuming the water reduces the stock, but a natural recharge (e.g. rain) can increase it. In forestry economics, the stock might be measured using the biomass of forestry, the number of tons of wood. Cutting down trees reduces the stock, but the forest's natural growth increases it. In climate economics, the stock might be measured using the parts per million (ppm) of atmospheric CO_2. Carbon emissions increase the stock, but some of the stock is absorbed into other carbon reservoirs, e.g. oceans. In all of these case, the stock might increase or decrease over time, depending on the relation between society's actions and the natural growth/recharge/decay.

13.1 Growth dynamics

Objectives and skills

- Understand the meaning of biomass and the growth function.
- Graph the logistic growth function and interpret its parameters.

For the sake of specificity, we consider fishery economics, where we measure the stock of fish using biomass, e.g. the number of tons of fish. Biomass does not capture the population age and size distribution: twenty half-pound fish and ten one-pound fish both contribute ten pounds of biomass. The age and size distributions are hard to measure and they increase model complexity. For the purpose of explaining the basic renewable resource model, a single stock variable, biomass, is adequate.

We denote the stock of fish in period t as x_t, so the change in the stock is $x_{t+1} - x_t$. The growth function, $F(x)$, describes the evolution of the fish stock in the absence of harvest. Some fish die and new fish are born, so the stock might increase or decrease over time. Growth depends on the stock:

$$x_{t+1} - x_t = F(x_t).$$

Growth also depends on the possibly random stocks of predators and prey and changes in pollution concentrations and ocean temperature. Even holding those features fixed, there may be intrinsic randomness of the growth of x. These types of considerations lead to a more descriptive but more complicated model, and we ignore them.

If a period equals one year, then $F(x)$ equals the annual growth, $\frac{F(x)}{x}$ equals the annual growth *rate* and $\frac{F(x)}{x}100$ equals the percent growth rate. The most common functional form for the growth function is the Shaeffer, or "logistic" model:

$$F(x_t) = \gamma x_t \left(1 - \frac{x_t}{K}\right). \tag{13.1}$$

This model uses two parameters, $\gamma > 0$, and $K > 0$. The parameter K is the "carrying capacity", measuring the level of stock that can be sustained, absent harvest. Growth is zero if $x_t = K$ or if $x_t = 0$; the stock grows if $0 < x_t < K$ and it falls if $x_t > K$. Congestion decreases the carrying capacity, K. As the stock increases, the fish compete for prey, and/or they become more vulnerable to predators. This congestion limits the potential growth of the stock.

The growth rate of the stock with the logistic growth function is

$$\frac{x_{t+1} - x_t}{x_t} = \frac{F(x_t)}{x_t} = \frac{\gamma x_t \left(1 - \frac{x_t}{K}\right)}{x_t} = \gamma \left(1 - \frac{x_t}{K}\right). \tag{13.2}$$

The parameter γ is the "intrinsic growth rate". In the absence of congestion ($x_t = 0$ or $K = \infty$), the growth rate equals γ. A larger value of K implies a higher growth rate (less congestion) for given x. For given γ and K, the growth rate falls with x, so the *growth rate* (*not* growth) reaches the maximum value, γ, at $x = 0$. The value $\gamma = 0.07$, for example, means that in the absence of congestion, the stock grows at 7% per year. For x close to 0, congestion is relatively unimportant, and the growth rate is close to 7%. However, as the stock gets larger, congestion becomes more important, until growth ceases as x approaches the carrying capacity, K. Figure 13.1 shows graphs of the logistic growth function for three different growth rates, and $K = 50$. For positive stocks, a larger γ implies larger growth.

13.2 Harvest and steady states

Objectives and skills

- Understand the meaning of and be able to graph "harvest rules".

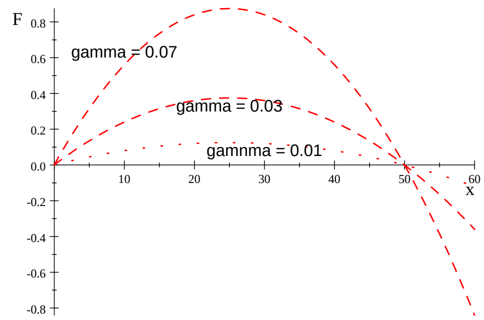

Figure 13.1: The logistic growth function for three values of γ.

- Understand the meaning of "steady states" and be able to identify them.

The introduction of harvest, $y > 0$, changes the dynamics. The amount harvested at a point in time might depend on the stock of biomass at that point in time. A "harvest rule" gives the harvest level as a function of the stock. Two examples illustrate harvest rules: $y(x) = \min(x, Y)$, where Y is a constant, and $y(x) = \mu x$, with $\mu > 0$ a constant. For the first example, harvest is constant at Y if this level is feasible, i.e. if $x \geq Y$. If the stock is less than Y, all of it is harvested. For the second example, harvest is a constant fraction of the stock. For example, for $\mu = 0.01$, annual harvest equals one percent of the fish stock. It is not possible to take more than the entire stock, so in the discrete time setting considered here, $\mu \leq 1$. Figure 13.2 shows the growth functions with the zero, constant, and the proportional harvest with $K = 50$, $\gamma = 0.03$, $Y = 0.1$, and $\mu = 0.01$. Under these two harvest rules, the change in the stock is

$$\begin{array}{rl} \text{constant harvest:} & x_{t+1} - x_t = \gamma x_t \left(1 - \frac{x_t}{K}\right) - \min(x_t, Y) \\ \text{harvest proportional to stock:} & x_{t+1} - x_t = \gamma x_t \left(1 - \frac{x_t}{K}\right) - \mu x_t. \end{array} \tag{13.3}$$

A steady state is any level of the stock at which growth minus harvest equals 0. A stock beginning at a steady state remains there. We have three examples of harvest rules: zero harvest, and the two rules shown in equation 13.3. We obtain the steady states, denoted x_∞, in these three cases by

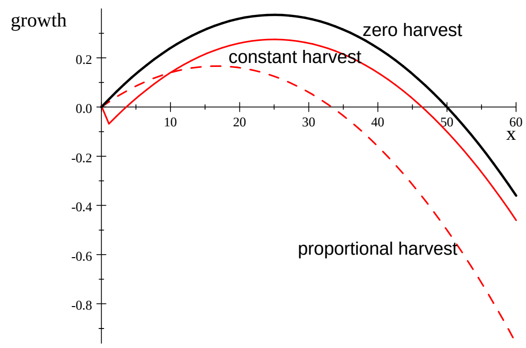

Figure 13.2: Growth with zero harvest (heavy solid curve), with harvest proportional to the stock (dashed), and with the stock equal to Y for $x \geq Y$ (light solid).

setting the growth minus harvest equal to 0 and solving for x:

$$\begin{aligned} y = 0\text{:} \quad & \gamma x \left(1 - \tfrac{x}{K}\right) = 0 \\ & \Rightarrow x_\infty, \in \{0, 50\} \\ y = \min(x, 0.1)\text{:} \quad & \gamma x \left(1 - \tfrac{x}{K}\right) - \min(x, 0.1) = 0 \\ & \Rightarrow x_\infty \in \{0, 3.6, 46.4\} \\ y = 0.01x\text{:} \quad & \gamma x \left(1 - \tfrac{x}{K}\right) - 0.01x = 0 \\ & \Rightarrow x_\infty \in \{0, 33\}. \end{aligned} \tag{13.4}$$

13.3 Stability

Objectives and skills

- Understand the distinction between stable and unstable steady states.
- Using graphical methods and *the continuous time model*, identify steady states and determine whether each is stable or unstable.
- Have an intuitive understanding of the relation between discrete and continuous time models, and understand the advantages of each.

A steady state is "stable" if the stock trajectory approaches that steady state when the stock begins sufficiently close to it. A steady state is "unstable" if the stock trajectory beginning close to it, moves away from it.

The stability or lack of stability of a steady state provides important information about the dynamics of the fish stock, and is therefore important in policy questions. Often there are multiple steady states, including $x = 0$. If this steady state is stable, a small stock eventually becomes extinct; if it is unstable, a small stock becomes larger over time.

13.3.1 Discrete time versus continuous time models

Discrete and continuous time models have different advantages. With the discrete time model, we can derive the necessary condition for optimality (the Euler equation) using only elementary calculus. The "no-intertemporal-arbitrage" interpretation of the Euler equation is also more intuitive in the discrete time setting. The continuous time model has three advantages. First, as noted in previous chapters, the graphs of the continuous time equilibrium trajectories are easier on the eye, because they are smooth instead of step functions. Second, some computations are easier in the continuous time setting. Third, the analysis of stability is *much* easier in the continuous time setting. At the cost of mathematical rigor, we take advantage of both the discrete and continuous time approaches. We use the discrete time setting to present and interpret the models and the Euler equation, but we use the continuous time analog to study stability. Here we explain the difficulty arising with discrete time stability analysis, and then discuss the relation between the two models (cf Appendix G).

In the discrete time setting there is a non-negligible change in variables (e.g. the stock of fish) from one period to the next, outside of a steady state. The stock might jump from one interval to another where the behavior is quite different. This possibility can lead to chaos, where paths (trajectories) are very irregular (they do not repeat in a finite amount of time) and very sensitive to the initial condition. It is possible to rule out chaotic behavior by restricting parameter values, but that still leaves special cases. A steady state might be stable (meaning that paths starting close to the steady state approach it), but the approach path might be monotonic (steadily increasing or decreasing over time) or cyclical (first increasing, then decreasing, then increasing, and so on). These possibilities are tangential to our concerns. Finally, in the discrete time setting we cannot determine the stability or instability of a steady state merely by examining a graph; we require calculation.

For the purpose of considering stability, we consider "the continuous time analog" to the discrete time model. For our one-dimensional models, the dynamics in continuous time are simple, and stability can be determined by inspection of a graph, without calculation. We develop the continuous time analog using the general growth function and harvest rule, $F(x)$ and $y(x)$, replacing equation 13.3 with the more more general relation

$$x_{t+1} - x_t = F(x_t) - y(x_t). \tag{13.5}$$

Instead of studying the stability of steady states of this equation, we study the stability of steady states of the continuous time analog, the ordinary differential equation[1]

$$\frac{dx_t}{dt} = F(x_t) - y(x_t). \tag{13.6}$$

These two equations have the same steady states, where $F(x) - y(x) = 0$.

Caveat One subtly arises in "moving" from discrete to continuous time. Suppose that we pick a unit of time equal to a year, so that in the discrete time setting y equals the amount harvested in a year. In this setting, we have the constraint $y \leq x$, because it is not possible to harvest more fish than the level of biomass. This constraint does not apply in the continuous time setting. An example helps to clarify this claim. We have a stock of wealth, \$1000, we cannot borrow, and we receive no interest on savings; our unit of time is a year. In the discrete time setting, y equals the amount spent in a year; because we cannot borrow, we cannot spend more than \$1000: $y \leq x = 1000$. In the continuous time setting, y equals spending per unit of time. Here, y can take any non-negative value. For example, it is feasible to spend \$1000 per year for the duration of a year. It is also feasible to spend \$5000 per year *for the duration of one-fifth of a year.* We can spend at any rate – just not for very long. In the discrete time setting we have to honor the constraint $y \leq x$, but in the continuous time setting we require only $x \geq 0$.

[1] Equations 13.5 and 13.6 have the same steady states: $x_{t+1} - x_t = 0$ if and only if $\frac{dx_t}{dt} = 0$. Provided that the length of a period in the discrete time setting is sufficiently small, the dynamics of the continuous and discrete systems are similar in the neighborhood of a steady state.

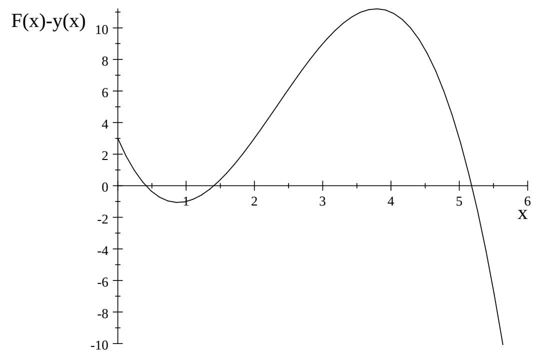

Figure 13.3: The graph of $\frac{dx}{dt} = F(x) - y(x)$. There are three steady states, at $x = 0.4$, $x = 1.4$ and $x = 5.2$. The first and third are stable, and the middle steady state is unstable

13.3.2 Stability in continuous time

It is simple to determine whether an interior steady state is stable in the continuous time setting. First, we identify the steady state(s) by finding the solution(s) to the equation $F(x) - y(x) = 0$, exactly as in the discrete time setting. Figure 13.3 shows the graph of an arbitrary function $F(x) - y(x)$ (one without a specific resource interpretation). This function has three roots, i.e. three steady states, where the graph crosses the x axis. Consider the low steady state, $x = 0.4$. We see that for a value of x close to but slightly below 0.4, $\frac{dx}{dt} = F(x) - y(x) > 0$, i.e. x is becoming larger over time. For a value of x close to but slightly above 0.4, $\frac{dx}{dt} = F(x) - y(x) < 0$, i.e. x is becoming smaller. Therefore, we conclude that $x = 0.4$ is a stable steady state: a trajectory beginning close to, but not equal to $x = 0.4$ approaches the level $x = 0.4$. A parallel argument shows that the middle steady state, $x = 1.4$, is an unstable steady state, and the large steady state, $x = 5.2$, is a stable steady state.

Noticing that the slope of $F(x) - y(x)$ is negative at the two stable steady states in Figure 13.3, and the slope is positive at the unstable steady state,

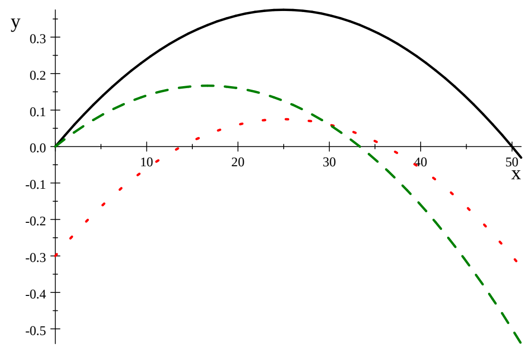

Figure 13.4: Solid curve: the graph of the logistic curve, $F(x)$. Dashed curve: the graph of $F(x) - 0.01x$. Dotted curve: the graph of $F(x) - 0.3$.

we obtain the following rule for checking stability.[2]

$$\begin{array}{c} x_\infty \text{ is a } \textit{stable} \text{ steady state if and only if } \frac{d(F(x_\infty)-y(x_\infty))}{dx} < 0 \\ x_\infty \text{ is an } \textit{unstable} \text{ steady state if and only if } \frac{d(F(x_\infty)-y(x_\infty))}{dx} > 0. \end{array} \tag{13.7}$$

We can determine the sign of these derivatives without calculation, merely by inspection of the graph of $F(x) - y(x)$.

13.3.3 Stability in the fishing model

Figure 13.4 shows the graphs of $F(x) - y(x)$ (the growth function minus the harvest function) using the logistic growth function $F(x)$ and three harvest rules: $y(x) = 0$ (solid); $y(x) = 0.01x$ (dashed); and $y(x) = 0.3$ (dotted) The rule in equation 13.7 tells us whether various steady states are stable or unstable. The points of intersection of the graphs in Figure 13.4 and the x axis are steady states. Absent harvest, the solid curve shows that the steady states are $x = 0$ and $x = K = 50$. Using the rule in equation 13.7, we see that $x = 0$ is an unstable steady state and $x = K$ is a stable steady state. The dashed curve shows that under proportional harvest, $y = 0.01x$,

[2] If $\frac{d(F(x_\infty)-y(x_\infty))}{dx} = 0$, points on one side of the steady state approach the steady state, and points on the other side move away from the steady state. We do not discuss this knife-edge case.

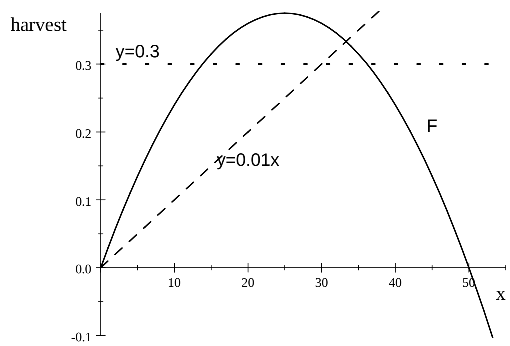

Figure 13.5: The logistic growth function, and two harvest rules: constant harvest, $y = 0.3$ (dotted) and harvest proportional to stock, $y = 0.01x$ (dotted).

the two steady states are $x = 0$ and $x = 33.33$. Again, the low steady state is unstable and the high steady state is stable.

The dotted curve, corresponding to constant harvest $y = 0.3$, has two points of intersection with the x axis: $x = 13.8$ and $x = 36.2$. The low steady state is unstable, and the high steady state is stable. Under constant harvest, $x = 0$ is also a *stable* steady state, even though the graph of $F(x) - 0.3$ does not intersect the x axis. Recall the "caveat" above. In the continuous time model, y can take any finite value. For our example, y can remain at 0.3 as long as $x > 0$. If stock is small, here lower than 13.8, constant harvest $y = 0.3$, exceeds natural growth, and the stock falls. As soon as the stock hits $x = 0$, harvest must stop. The stock heads to extinction, $x = 0$.

Figure 13.4 illustrates an important possibility that we will encounter again. Under zero harvest or harvest proportional to the stock, the stock always approaches the high steady state ($x = 50$ and $x = 33.33$ in the two examples), provided that the initial stock is positive. In contrast, under constant harvest, beginning with a positive stock, the stock might eventually approach either stable steady state, $x = 0$ or $x = 36.2$. The unstable steady state, $x = 13.8$, is a critical stock level. For initial stocks above this critical level, the stock approaches the high steady state, and for initial stocks below this level, the stock approaches the low steady state, 0.

A different perspective Figure 13.5 shows graphs of the growth and the harvest functions in the same figure, whereas Figure 13.4 shows their difference. We can use either figure to identify steady states and their stability. The reader should confirm that, as we move from left to right in Figure 13.5:

> (i) If the harvest function cuts the growth function from above, the associated steady state is unstable.
>
> (ii) if the harvest function cuts the growth function from below, the associated steady state is stable.

For example, under constant harvest $y = 0.3$, harvest exceeds growth for x below the low (interior) steady state. At these stock levels, the stock is falling: the stock moves away from the low steady state, so that steady state is unstable. In contrast, for stocks between the two interior steady states, growth exceeds harvest, so the stock moves toward the high steady state. Similarly, for stocks above the high steady state, harvest exceeds growth, so the stock declines, toward the high steady state.

13.4 Maximum sustainable yield

Objectives and skills

- Know the definition of Maximum Sustainable Yield (MSY) and be able to calculate the MSY for the logistic growth function.
- Understand the economic factors that determine whether optimal steady state harvest should be above or below MSY.

The maximum sustainable yield is the largest harvest that can be sustained in perpetuity. *Any* point on the graph of the growth function is a sustainable harvest. We can pick any point on this graph and draw a second graph intersecting that point; this second graph is a particular harvest rule for which the chosen point is a steady state, and thus a sustainable harvest. The *maximum sustainable harvest* occurs at the highest point on the graph of the growth function. We identify this highest point by solving $\frac{dF(x)}{dx} = 0$, the first order condition for the problem of maximizing $F(x)$. For the logistic growth function, this condition is

$$\frac{d\left(\gamma x\left(1-\frac{x}{K}\right)\right)}{dx} = \frac{1}{K}\gamma\left(K-2x\right) \overset{\text{set}}{=} 0.$$

Solving this equation gives the steady state stock corresponding to maximum sustainable yield

$$x = \frac{K}{2}.$$

Substituting this value into the growth function, $\gamma x \left(1 - \frac{x}{K}\right)$, gives the maximum sustainable yield

$$\text{MSY:} \;\; y = \frac{1}{4} K\gamma.$$

(The maximum sustainable *yield* is a level of y, not a level of x.) An increase in either the intrinsic growth rate, γ, or the carrying capacity, K increases the maximum sustainable yield.

The socially optimal steady state

Chapter 15 examines the socially optimal steady state, but even at this stage we can use economic logic to identify the factors that determine its level. We need a criterion for comparing different outcomes. The most common criterion (discounted utilitarianism) is the present discounted value of the stream of consumer and producer surplus. What steady state is optimal under this criterion? By definition, steady state output = consumption is highest at the MSY, so consumer surplus is highest there. Producer surplus, equal to revenue minus costs, depends on the cost of catching fish. If it is cheaper to catch fish when there are many fish (the stock is large), increasing x above $\frac{K}{2}$ reduces costs. With stock dependent harvest costs, the increase in the stock and the consequent decrease in the costs might increase producer surplus. At this level of generality we do not know whether the change actually increases producer surplus, because the change also alters revenue.

Even if higher stocks lower harvest costs, we do not know (at this level of generality) whether it is socially optimal to have a steady state stock above $\frac{K}{2}$. With a positive discount rate, a future benefit (e.g. a higher sum of consumer and producer surplus) is less valuable than a current benefit. Therefore, a positive discount rate encourages society to consume more today, leaving less for the future, and reducing the stock below the MSY.

In summary, consideration only of the consumption benefit suggests that the MSY is the optimal steady harvest. Recognition that harvest costs (might) depend on stock size suggests that the optimal steady state stock might lie above the level of MSY. Taking into account that the future is worth less than the present (due to discounting) suggests that society should aim for a steady state stock below the level of MSY.

13.5 Summary

We measure the stock of fish as biomass, e.g. the number of tons of fish. The growth equation determines the stock in the subsequent period as a function of the current stock. The logistic growth function depends on two parameters, the intrinsic growth rate γ and the carrying capacity K. A harvest rule, a function of the stock of fish, determines the harvest in a period. We emphasized two harvest rules, one equal to a constant, and the other equal to a constant fraction of stock.

At a steady state, the fish stock remains constant over time. The steady state depends on both the growth function and the harvest rule. There may be multiple steady states. A steady state is stable if and only if stocks that begin sufficiently close to the steady state converge to that steady state. If a trajectory beginning at any initial condition close to but not equal to the steady state moves away from that steady state, the steady state is unstable.

We introduced the continuous time model, for the purpose of making it easy to determine the stability or instability of a steady state. In order to relate the discrete and the continuos time models, the reader should think of the length of a period in our discrete time setting as being very small. If the growth function is $F(x)$ and the harvest function $y(x)$, then $\frac{dx}{dt} = F(x) - y(x)$. Any solution to $F(x) - y(x) = 0$ is a steady state. The slope of $F(x) - y(x)$ is negative at a stable steady state and positive at an unstable steady state. We also need to consider levels of the state at which an inequality constraint binds. In the fishing context, the biomass cannot be negative, so $x \geq 0$; $x = 0$ is a steady state, which might be either stable or unstable, depending on the relation between the growth and harvest functions.

The maximum sustainable yield equals the maximum point on the growth function. For the logistic growth model, the maximum sustainable yield occurs where the stock is half of its carrying capacity.

13.6 Terms, study questions, and exercises

Terms and concepts

Biomass, stock variables, annual growth rate, logistic growth (or logistic model), carrying capacity, intrinsic growth rate, harvest rule, steady state, chaos, stability, monotonic path, cyclical path, Maximum Sustainable Yield.

Study questions

1. Given the graph of a growth function $x_{t+1} - x_t = F(x)$, you should be able to identify the carrying capacity and the maximum sustainable yield, and say in a few words what each of these mean.

2. For a differential equation $\frac{dx}{dt} = G(x)$, if you are shown a graph of $G(x)$ you should be able to identify the steady state(s) and say which, if any of these are stable. You should be able to explain your answer in a couple of sentences.

3. (a) Given a single figure that shows both the graph of the growth function $F(x)$ and the harvest function $y(x)$, should be able to identify the steady states and explain (in very few words) which is stable and which is unstable. You should be able to sketch a graph of their difference, $\frac{dx}{dt} = F(x) - y(x)$, and use the rule in equation 13.7 to confirm that your answer to part (a) was correct.

Exercises

1. Using an argument that parallels the discussion of the low steady state in Figure 13.3, explain why the middle steady state is unstable and why the high steady state is stable.

2. For the logistic growth function, $F(x) = \gamma\left(x - \frac{x}{K}\right)$, identify the proportional harvest rule (the value of μ in the rule $y = \mu x$) that supports the maximum sustainable yield as a steady state. Is this steady state stable?

3. The "skewed logistic" growth function is

$$F(x_t) = \gamma x_t \left(1 - \frac{x_t}{K}\right)^{\phi},$$

with $\phi > 0$. For $\phi = 1$ we have the logistic growth function in equation 13.1. (a) How does the magnitude of ϕ affect the growth rate? (b) The maximum sustainable yield occurs at $x = \frac{K}{\phi+1}$. Derive this formula. (c) Use a software package of your choice to draw this curve for $K = 50$, $\gamma = 0.03$, and for both $\phi = 2$ and $\phi = 0.5$.

Sources

Clark (1996) is the classic text on renewable resource economics, and fishery economics in particular.

Hartwick and Olewiler (1986) cover much of the material in this chapter.

Conrad (2010) presents the discrete time material.

Readers interested seeing how the deterministic models discussed in this book can be extended to the stochastic setting should consult Mangel (1995).

Chapter 14

The open access fishery

Objectives

- Analyze policy under open access.

Skills

- Use the zero-profit condition to obtain the open access "harvest rule".
- Determine the evolution of biomass under open access.
- Understand how a tax affects harvest incentives and the evolution of biomass.

A tax changes the level of harvest for a given level of the stock, and it changes the evolution of the stock. If there are property rights, resource owners take into account future costs and benefits in making current extraction decisions. Absent property rights individuals have no reason to think about the (negligible) effect their harvest has on future stocks. Here, agents choose their current harvest to maximize their *current* profit.

Chapter 12.3.1 studies a static version of this scenario, where everything happens in the same period. Dynamics are central to the fishery problem, where the externality unfolds over time. Here we study a model in which fishers' aggregate current harvest affects the subsequent stock, resulting in a dynamic externality. Chapter 12.3 points out that the first best harvest typically involves many types of decisions; regulators rarely have as many instruments (policy variables) as there are targets. We show how a tax on catch, known as a landing fee, influences the open access equilibrium.

14.1 Harvest rules

Objectives and skills

- Given an inverse demand function and harvest cost function, find the open access harvest rule.
- Sketch the graphs corresponding to this harvest rule and the logistic growth function.
- Determine the steady states and their stability, and answer a comparative statics question.

We consider two cases, the first with constant stock-independent average and marginal harvest costs, $c(x,y) = Cy$, and the second with stock dependent average and marginal cost, $c(x,y) = C\frac{y}{x}$. Due to free entry and exit, there are zero profits at every point in time, so price equal average cost. If profits were positive, new entrants would increase supply, lowering the price and lowering profits; if profits were negative, current fishers would leave the industry, lowering supply, increasing the price and increasing profits.

14.1.1 Stock-independent costs

With stock-independent constant harvest costs, $c(x,y) = Cy$, "price equals average cost" requires $p_t = C$. For the inverse demand function $p = a - by$, this condition implies $y = \frac{a-C}{b}$. For $C < a$, open access harvest is positive whenever the stock is positive; the open access harvest rule is[1]

$$y(x) = \min\left(x, \frac{a-C}{b}\right) \tag{14.1}$$

Chapter 13 examines the dynamics under this kind of harvest rule. Figure 14.1 reviews this material, showing the graph of a logistic growth function and the graphs of two harvest rules, corresponding to a low and a high cost, C (the dashed and dotted lines, respectively). For both of these costs, there are three steady states. The zero steady state and the high steady state (occurring to the right of the maximum sustainable yield) are both stable. The middle steady state is unstable.

[1] Equation 14.1 is the harvest rule in the discrete time setting. In the continuous time setting, were the rate y can be arbitrarily large, the harvest rule is $y = \frac{a-C}{b}$ for $x > 0$ and $y = 0$ for $x = 0$.

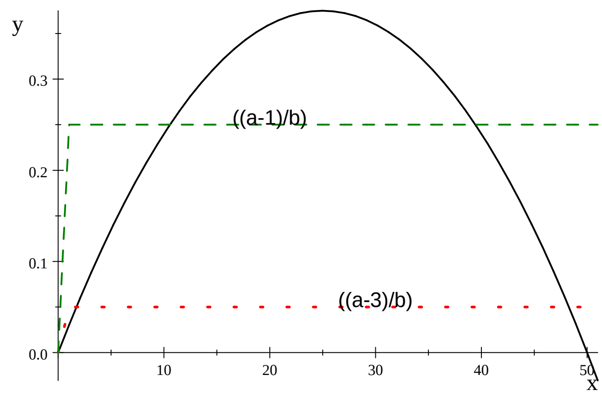

Figure 14.1: The solid curve is the graph of the logistic growth function with $\gamma = 0.03$ and $K = 50$. The dashed graph shows the open access harvest rule for $C = 1$ and the dotted graph shows the harvest rule for $C = 3$. Inverse demand is $p = a = by$ with $a = 3.5$ and $b = 10$.

14.1.2 Stock-dependent costs

We provide a "micro-foundation" for the cost function $c(x, y) = C\frac{y}{x}$, and then derive the open access harvest rule.

Micro-foundation of the cost function Fishing "effort", E, is an amalgam of all of the inputs in the fishery sector. In a representative agent model, E_t is the *aggregate* effort in the fishery in period t. Greater effort increases harvest. Fish are easier to catch when the stock is large, so for a given amount of effort, a larger stock increases harvest. The fishery production function shows how effort, E, and the stock, x, determine harvest, y. The simplest production function assumes that the level of harvest per unit of effort is proportional to the size of the stock, $\frac{y}{E} = qx$, or

$$y = qEx \Rightarrow E = \frac{y}{qx}. \tag{14.2}$$

The parameter $q > 0$ is the "catchability coefficient". A larger q means that for a given stock size, fishers need less effort to obtain a given level of harvest. The cost per unit of effort is the constant, w. If one "unit of effort" equals one boat and 200 hours of labor and a particular net, then w equals the cost of renting the boat, paying the crew, and buying or renting the net.

The cost of harvesting y, given the stock x, is

$$\text{harvest cost: } c(x,y) = wE = \frac{w}{qx}y = \frac{C}{x}y, \tag{14.3}$$

where $C \equiv \frac{w}{q}$. Harvest costs fall as: the stock of fish (x) rises, the cost per unit of effort (w) falls, and the catchability coefficient (q) increases.

The harvest rule If harvest is positive in the open access equilibrium, then price equals average costs. With linear inverse demand, $p(y) = a - by$, price equals average cost requires $a - by = \frac{C}{x}$. If cost exceeds the choke price, a, then harvest equals zero. These two facts imply the open access harvest rule[2]

$$y(x) = \begin{cases} \frac{1}{b}\left(a - \frac{C}{x}\right) \text{ for } \frac{1}{b}\left(a - \frac{C}{x}\right) \geq 0 \\ 0 \text{ for } \frac{1}{b}\left(a - \frac{C}{x}\right) < 0. \end{cases} \tag{14.4}$$

Figure 14.2 shows the graph of the logistic growth function and the open access harvesting rule in equation 14.4 for "low demand", $p = 3.5 - 10y$ (dashed) and "high demand", $p = 4.2 - 10y$ (solid) with $C = 5$. The figure identifies the three interior steady states (where $x > 0$), points A, B, D, under low demand. In the high demand scenario, the only interior steady state occurs at a low stock level, close to but slightly lower than point A. Extinction, $x = 0$, is a steady state in both cases.

In the *low demand* scenario, moving from left to right, the dashed curve cuts the growth function from below at points A and D, and it cuts the growth function from above at point B. Thus, points A and D are stable steady states, and point B is an unstable steady state (cf. the final paragraph in Section 13.3.3). The point $x = 0$ is an unstable equilibrium. For sufficiently small but positive stock, $a - \frac{C}{x} < 0$, so $y = 0$. Here, the stock is so low, and the harvest costs so high, that the equilibrium harvest is 0. Because growth is positive for small positive x, the fish stock is growing in this region. Therefore, if the initial stock is positive and below point B, the stock in the open access fishery converges to point A. If the initial stock is above point B, the stock converges to point D.

In the *high demand* scenario (dotted curve), there are two steady states. The stable steady state is slightly below point A; $x = 0$ is an unstable

[2] A necessary and sufficient condition for $y(x) < x$ for all $x \geq 0$ is $a < 2\sqrt{bC}$. When this inequality holds, 14.4 is the harvest rule for both the discrete and continuous time setting.

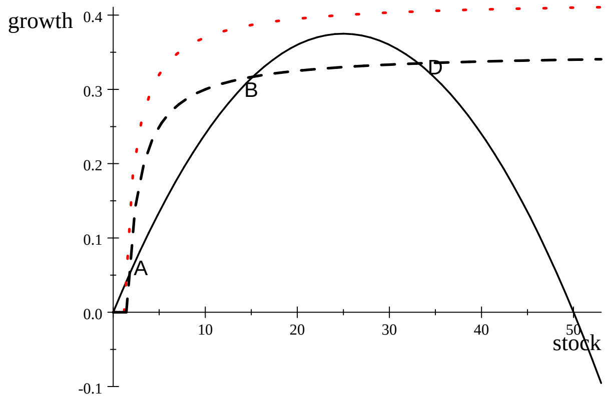

Figure 14.2: The solid curve shows the logistic growth function with $\gamma = 0.03$ and $K = 50$. Dashed curve: harvest rule for $a = 3.5$. Dotted curve: harvest rule for $a = 4.20$.

steady state. Here, for any positive initial stock, the stock under open access converges to a level slightly below point A.[3]

14.2 Policy applications

Objectives and skills

- Understand the effect of a constant tax on the evolution of the biomass and on the steady states.
- Understand why the long run effect of the tax may depend on the level of the stock at the time the tax is imposed.

Policy affects the equilibrium outcome by altering the equilibrium harvest rule. Gear restrictions increase C, shifting down the harvest rule (cf. equations 14.1 and 14.4). Here we consider the effect of taxes when average harvest costs depend on the stock.

Chapter 10 shows that (in a closed economy) a tax has the same effect regardless of whether consumers or producers have statutory responsibility for paying it. If consumers face price p and have inverse demand $p = a - by$,

[3] Appendix H provides a different way to visualize the equilibrium, using the concept of a "bioeconomic equilibrium".

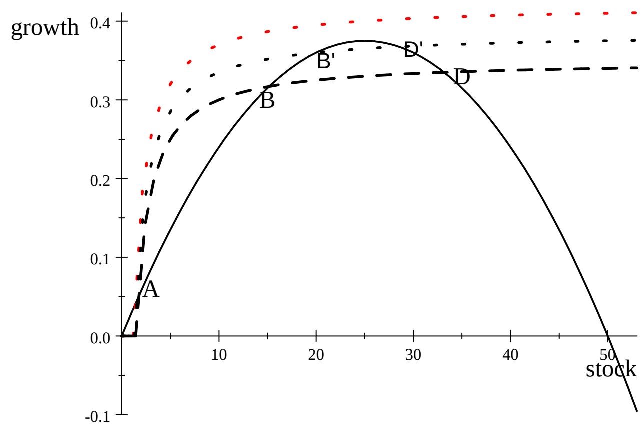

Figure 14.3: The logistic growth function (the solid graph) and harvest rules for:: zero tax (highest graph); $\nu = 0.35$ (middle graph); and $\nu = 0.7$ (lowest graph). Demand is $p = 4.2 - 10y$.

producers facing the unit tax ν receive the net-of-tax price $p - \nu$. Zero profits requires that this price equals average cost, or $a - by - \nu = \frac{C}{y}$. Modifying equation 14.4, the open access harvest rule under the tax is

$$y(x) = \begin{cases} \frac{1}{b}\left(a - \nu - \frac{C}{x}\right) \text{ for } a - \nu - \frac{C}{x} \geq 0 \\ 0 \text{ for } a - \nu - \frac{C}{x} < 0. \end{cases} \quad (14.5)$$

A one unit decrease in a or a one unit increase in ν have the same effect on $a - \nu$, and thus have the same effect on the harvest rule. Figure 14.3 shows harvest rules for $a = 4.2$, and three values of the tax: $\nu = 0$; $\nu = 0.35$ (so $a - \nu = 3.85$); and $\nu = 0.7$ (so $a - \nu = 3.5$).

The role of the initial condition The "initial condition" is the level of the stock, x_0, at the time the tax is first imposed. The long run (steady state) effect of the tax might depend on the initial condition. The low steady states for the three harvest rules (corresponding to the three tax levels) are slightly different, but they all occur so close to point A, where $x_A = 1.7$, as to be indistinguishable in Figure 14.3. However, the intermediate and the high steady states are appreciably different under the two positive taxes; these steady states do not exist if $\nu = 0$. The points B' and D' are steady states under the tax $\nu = 0.35$, and B and D are steady states under the tax

$\nu = 0.7$. The stocks at these points are

$$x_A = 1.7, \;\; x_B = 15.4, \; x_{B'} = 20, \;\; x_{D'} = 28.6, \text{ and } x_D = 33.2.$$

We do not distinguish between x_A and $x_{A'}$ because these are so close that their difference is not interesting.

Figure 14.3 shows that the initial stock, – not just the policy level – can have a significant effect on the steady state to which the stock converges. If $x_0 < 15.4$, neither tax has an appreciable effect on the steady state: the stock converges to a point close to $x_A = 1.7$ for $\nu \in \{0, 0.35.0.7\}$. If $15.4 < x_0 < 20$, the tax $\nu = 0.35$ has a negligible effect on the steady state (close to $x_A = 1.7$), but the tax $\nu = 0.7$ causes the stock to converge to 33.2. If $x_0 > 20$, then the stock converges to $x = 28.6$ for $\nu = 0.35$ and to $x = 32.2$ for $\nu = 0.7$. In this example, if the initial stock is "moderately small" ($15.4 < x_0 < 20$), then $\nu = 0.35$ has a negligible effect on the steady state, whereas the tax $\nu = 0.7$ leads to a large increase in the steady state. If the initial stock is "moderately large", ($x_0 > 20$) then both taxes lead to qualitatively different steady states, compared to $\nu = 0$. Under the zero tax, the stock converges to the low steady state stock, for any positive x_0.

Steady state welfare effects of the tax Here we assume that $x_0 > 20$. The low tax leads to a lower steady state stock, but a higher steady state harvest, compared to the high tax: point D' is to the left and above point D. A higher harvest corresponds to a lower consumer price, and higher consumer surplus. In the high steady state under the high tax, $x_D = 33.2$, so harvest = consumption equals

$$y = \gamma x_D \left(1 - \frac{x_D}{K}\right) = 0.335.$$

Consumers are willing to purchase this amount if they face the tax-inclusive price $4.2 - 10 \times 0.335 = 0.853$. At this price, consumer surplus is

$$\int_{0.853}^{4.2} \left(\frac{4.2 - z}{10}\right) dz = 0.56$$

and tax revenue is $\nu y = 0.7\,(0.335) = 0.234\,5$. Profits are zero at every point, so social welfare equals the sum of consumer surplus plus tax revenue, 0.794. Table 1 collects these numbers at the high stable steady states under the two taxes. It shows that steady state consumer surplus is higher, tax revenue is lower, and their sum, social welfare, is higher under the smaller tax.

tax (ν)	stock	harvest	price	consumer surplus	tax revenue	social welfare
0.7	33.2	0.335	0.853	0.56	0.234	0.794
0.35	28.6	0.367	0.528	0.67	0.129	0.803

Table 14.1: *High stable* steady state stock, harvest, price, consumer surplus, tax revenue, and social surplus for two taxes.

Policy implications Even if a tax has only a negligible effect on the steady state to which the stock converges, it may nevertheless have a significant effect on welfare, by slowing the decline of the fishery. For example, if x_0 is slightly lower than 15.4, the tax $\nu = 0.35$ causes a 9% reduction in the initial harvest (relative to harvest under $\nu = 0$) and the tax $\nu = 0.7$ causes an 18% reduction in the initial harvest. This reduction in harvest is not large enough to keep the stock from converging to approximately the same low level, but the higher tax slows the fishery's decline. Therefore, for an initial condition below 15.4, the present discounted stream of welfare is likely higher under the larger tax.

The effect of the tax on the steady state depends on the initial condition and on model parameters. This kind of model provides only a rough guide for policy, because it only approximates the real world, and because data limitations make it hard to estimate the model parameters and the current stock level. The model does, however, reveal some trade-offs. In view of the amount of uncertainty in going from the real world to the model, the regulator might want to build in a margin of safety, choosing a tax that exceeds the optimal level implied by the model. The larger tax protects against the possibility that we over-estimated the initial biomass, or were wrong about some other key parameter.

Thus far we have considered only a constant tax, a number. That single number can be used to target (i.e., to select) a single endogenous variable. We took the steady state stock, and corresponding payoff, as the target of interest. However, it makes sense for the regulator to care about the payoff along the trajectory en route to the steady state. A richer description of the regulatory problem includes "state-contingent" taxes, defined as taxes that vary with the level of the stock (cf. Chapter 15.2).

14.3 Summary

In an open access equilibrium with costless entry and exit, profits are zero, implying that price equals average harvest cost. We obtained a closed form expression for the open access harvest rule under linear demand and two types of cost function. Using these harvest rules and the logistic growth function, we identified the steady states and their (in-)stability, and studied the evolution of the stock of fish under open access. For the case of stock-dependent harvest costs, there might be either a single interior (= positive) steady state or three interior steady states. In the former case, the unique interior steady state is stable. In the latter case, the low and the high interior steady states are stable, and the middle steady state is unstable. The stock $x = 0$ is an unstable steady state under stock-dependent harvest cost, and it is a stable steady with constant average harvest costs.

A unit tax on harvest shifts down the open access harvest rule, reducing harvest for any level of the stock. Under stock-dependent harvest cost, a sufficiently high unit tax moves the fishery from the situation where there is a single interior steady state to the situation where there are three interior steady states. Here, the tax creates a stable steady state with a high level of the stock. The effect of a tax depends on both the magnitude of the tax and on the level of the stock at the time the tax is first imposed (the initial condition). We also used this example to determine the steady state level of welfare (= consumer surplus plus tax revenue) under different taxes.

14.4 Terms, study questions, and exercises

Terms and concepts

Myopic, catchability coefficient, initial condition, steady state supply function.

Study questions

1. (a) For the case of constant average harvest cost, Cy, linear inverse demand, $p = a - by$, and logistic growth, obtain the open access harvest rule, and sketch it on the same figure as the growth function. Illustrate how a change in C alters the steady state and the dynamics of the fish stock. (b) Illustrate and explain the effect of a constant unit tax, ν,

on the steady states. (c) Use a graph and short discussion to illustrate the fact that the effect of a given tax depends both on its magnitude and on the stock level at the time the tax is imposed.

2. Begin with the production function showing output (harvest) $y = qEx$, where E is effort and x is biomass. Explain the meaning of E (perhaps by using an example) and the meaning of the parameter q. What is the name given to this parameter?

3. Suppose that a unit of effort costs w. Derive the cost function for the example in the previous question. Explain your derivation. (It is not enough to merely memorize this cost function.)

4. Using the cost function in the previous question and the inverse demand curve $p = a - by$, derive the open access harvest rule. Be able to explain each step. (Memorization is not sufficient.)

5. Write down the logistic growth function. On the same figure, sketch graphs of this growth function and the harvest rule obtained in the previous question.

6. Using the figure from the previous question, show the effect of an increase in the demand slope, b, or intercept, a, on harvest rule, and on any steady state(s). Discuss the effect of this change (in a parameter of the demand function) on the *evolution* of the stock. A complete answer must explain how the parameter change can qualitatively alter the evolution of the stock, depending on the initial condition.

7. Show how a constant tax affects the open access harvest rule.

8. By means of a graph, show how an increase in the tax can alter the steady state(s).

9. Explain why the effect of the tax depends on both the magnitude of the tax and on the initial condition of the biomass, at the time the tax is first implemented.

Exercises

1. This exercise illustrates the fact that imposing restrictions typically increases production costs. Suppose that effort depends on the size

of the boat, B, and the amount of labor, L in the following manner: $E = B^{0.5}L^{0.5}$. The cost of the boat, bB is proportional to its size, with a "price per unit size b". The wage is w, so the labor bill is wL. Thus, the cost of providing one unit of effort is C, with

$$c = \min_{B,L} (bB + wL) \text{ subject to } B^{\beta}L^{1-\beta} = 1.$$

(a) Find the optimal labor/boat size ratio, $\left(\frac{L}{B}\right)^*$ and the cost of providing one unit of effort,, c, as a function of prices b and w (b) Suppose that regulation doubles the required labor/boat size ratio to $2\left(\frac{L}{B}\right)^*$. Denote as $\tilde{c}$, the cost of providing a unit of effort under this regulation, a function of b and w. Compare $\tilde{c}$ and c.

2. (a) Consider an open access fishery with constant harvest costs, $c(x, y) = Cy$. Use linear demand $p = a - by$ and the logistic growth function. Suppose that producers have the statutory obligation to pay a unit tax, ν. Obtain the open access harvest rule in this case. (b) What is the tax incidence in this model? (c) Does your answer to part b mean that the tax makes consumers worse off? (d) Create a figure with graphs of the logistic growth function and also showing two open access harvest rules, for $\nu = 0$ and for $\nu > 0$. (The precise positive value is unimportant. The point is to understand the qualitative properties of this figure.) Label the steady states under both harvest rules and identify which of these (if any) are stable. (e) Suppose that the regulatory regime switches from $\nu = 0$ to $\nu > 0$ when the stock of fish is between the two middle steady states (i.e., the middle steady state under $\nu = 0$ and the middle steady state under $\nu > 0$.) Describe the evolution of the fish stock under each regulatory regime.

3. Using Figure 14.2, show how a larger value of C alters the two harvest rules (corresponding to low demand and high demand). Using this information, describe the effect of a larger C on the steady state(s) in the low and high demand scenarios.

4. Using equation 14.4, provide the economic explanation for the statement that $y(x) = 0$ for $a - \frac{C}{x} < 0$.

5. Suppose that inverse demand is $p = a - \frac{a-1}{0.2}y$, with $a > 1$. (a) Show that for this demand function, the elasticity of demand, evaluated at

$p = 1$, is $\eta = \frac{1}{a-1}$. A larger value of a implies less elastic demand. (b) Using Figure H.1, show how the set of steady states changes as demand becomes more elastic (a decreases toward 1). Provide an economic explanation for your observation.

6. (*) Suppose that the production function for harvest is $y = (qEx)^{\phi}$ with $0 < \phi < 1$. The cost of a unit of effort is c, a constant. (a) Write the cost function (expressing cost of harvest as a function of harvest and the stock). Sketch the cost function as a function of harvest (for a given stock). (b) There is free entry, so that at each point in time, profits equal 0. Use this equilibrium condition, and the linear inverse demand function, $p = a - by$, to write an equation that gives harvest as an *implicit* function of the stock. (c) Although it is not possible to solve this equation to obtain harvest as an explicit function of the stock, it is simple to solve it to obtain the stock as an explicit function of the harvest. Using this approach, graph the relation between the harvest and the stock. It helps to pick a particular value of ϕ, e.g. $\phi = 0.5$. On the same graph, sketch the graph of the harvest rule when $\phi = 1$ (shown in Figure 14.2). (d) Transfer these two graphs onto a figure that also shows the growth function F, and use the result to describe the qualitative effect on the dynamics, of a decrease in ϕ (here, from 1 to 0.5). Provide an economic explanation. Hint, part a. Mimic the procedure used to find the cost function in the text. Rewrite the production function to find the level of effort needed to produce a given harvest, y, at a particular stock, x. How much does this level of effort cost? The answer is the cost function.

Sources In the model that we present, the size of the fishing fleet adjusts instantaneously, so profits equal zero at every moment. Conrad (2010) discusses extensions in which the industry's speed of adjustment depends on current profits, so profits can be non-zero outside a steady state.

Berck and Perloff consider a model with costly adjustment and rational expectations: firms' entry and exit decisions depend on their expectations of future profits.

Chapter 15

The sole-owner fishery

Objectives

- Understand the equilibrium condition in a price-taking sole owner fishery.

Skills

- Adapt the skills developed from the nonrenewable to the renewable resource setting.
- Write the sole owner's objective function and apply the perturbation method.
- Understand the role of growth in the no-intertemporal-arbitrage condition.
- Use the definition of rent to rewrite and re-interpret the Euler equation.
- Describe optimal policy in the presence of market failures.
- Identify the steady state for the optimally controlled fishery.

This chapter studies the price-taking sole-owner fishery, extending earlier results on *nonrenewable* resources. Absent other market failures, the sole owner harvests efficiently. By maximizing the present discounted stream of profits, she also maximizes the present discounted sum of producer and consumer surplus. There are few if any important sole owner fisheries. However, the sole owner outcome provides a baseline for determining policy when property rights are imperfect. We also consider the situation where the fishery provides ecological services that the sole owner does not internalize.

The tools developed to study the sole owner fishery are useful in many renewable resource settings For example, excessive greenhouse gas emissions occur because of the absence of property rights for the atmosphere. For both the open access fishery and the climate problem, the outcome under a social planner (or sole owner), provides information on optimal regulation.

We state the sole owner's optimization problem and then discuss the optimality condition, the Euler equation. The definition of rent leads to a more concise statement of this equation. We compare the sole owner and the open access steady states.

15.1 The Euler equation for the sole owner

Objectives and skills:

- State the sole owner's optimization problem.
- Write and interpret the Euler equation, and then express it using the definition of rent.

We consider two specializations of the parametric cost function. In the first, average harvest costs are constant in harvest and independent of the stock, $c(x,y) = Cy$. In the second, average harvest costs are constant in the harvest and decreasing in the stock $c(x,y) = \frac{C}{x}y$, so $\frac{\partial c(x,y)}{\partial x} = -\frac{C}{x^2}y < 0$. The owner of the resource takes the sequence of prices, p_0, p_1, p_2... as given; these prices are "endogenous to the model", via the inverse demand function, $p_t = p(y_t)$, but the resource owner takes them as exogenous.

For the constant average cost specification, the owner chooses the sequence of harvests, y_0, y_1, y_2... to solve

$$\max \sum_{t=0}^{\infty} \rho^t (p_t - C) y_t \tag{15.1}$$

For the stock dependent average cost specification, the owner solves

$$\max \sum_{t=0}^{\infty} \rho^t \left(p_t - \frac{C}{x_t}\right) y_t \tag{15.2}$$

In both cases, the owner faces the constraint $x_{t+1} = x_t + F(x_t) - y_t$.

For the sole owner facing constant harvest costs, C, the Euler equation is

$$p_t - C = \rho \left[(p_{t+1} - C) \left(1 + \frac{dF(x_{t+1})}{dx_{t+1}} \right) \right]. \tag{15.3}$$

The Euler equation for stock dependent case is (Appendix I)

$$p_t - \frac{C}{x_t} = \rho \left[\underline{\left(p_{t+1} - \frac{C}{x_{t+1}} \right) \left(1 + \frac{dF(x_{t+1})}{dx_{t+1}} \right)} + \underbrace{\frac{C}{x_{t+1}^2} y_{t+1}} \right]. \tag{15.4}$$

15.1.1 Intuition for the Euler equation

We emphasize the case of stock dependent harvest costs, equation 15.4. With one important difference, this necessary condition is identical to the Euler equation 5.2 for the nonrenewable resource. The right side of equation 15.4 involves

$$\left(p_{t+1} - \frac{C}{x_{t+1}} \right) \left(1 + \frac{dF(x_{t+1})}{dx_{t+1}} \right),$$

whereas the corresponding term with nonrenewable resources is

$$\left(p_{t+1} - \frac{C}{x_{t+1}} \right) (1 + 0).$$

These two expressions differ unless $\frac{dF(x_{t+1})}{dx_{t+1}} = 0$, i.e. unless the stock has no effect on growth. Stock-dependent growth is important in the renewable resource setting.

A trajectory consists of a sequence of harvest and stock levels. Along an optimal trajectory, a small change in harvest in some period, and an "offsetting change" in some other period, must lead to a zero first order change in the payoff: a perturbation does not improve the outcome. We obtain the Euler equation by considering a perturbation that changes harvest by a small amount in one period, and then makes an offsetting change in the next period, to return the stock to the candidate trajectory. The left side of equation 15.4 equals the marginal benefit of increasing harvest in period t, and the right side equals the marginal cost of the offsetting change in the subsequent period. Growth affects the change needed in period $t+1$ to offset the change in period t.

A savings example for intuition To provide intuition, we consider a savings problem unrelated to resources. An investor earns a per period return of r: investing a dollar at the beginning of a period produces $1 + r$ dollars at the end of the period. This investor has a "candidate savings plan", a trajectory of savings and wealth. The savings decision corresponds to harvest in the fishery setting, and wealth corresponds to biomass.

Suppose that the investor considers perturbing this candidate in period t by saving one dollar less than the candidate prescribes. In order to put her plan back on the candidate trajectory by period $t + 2$, she has to invest an additional $1 + r$ dollars in period $t + 1$, over and above the amount that her original ("pre-perturbation") plan calls for. The extra \$1 makes up for the dollar that she took out in period t, and the extra \$$r$ makes up for the interest that she lost by taking out that dollar. The same consideration applies in the fishery setting. In the savings problem growth in wealth (W_t) is a linear function of wealth (with slope equal to $1 + r$): Absent additional savings, wealth in the next period is $W_{t+1} = (1 + r) W_t$. In the fishery problem, growth in biomass is a nonlinear function of biomass.

Using this intuition Harvesting an extra unit in period t generates p_t additional units of revenue, and $\frac{C}{x_t}$ additional units of cost, for a net increase in profits of $p_t - \frac{C}{x_t}$, the left side of equation 15.4. This perturbation leads to lower and more expensive harvest in $t+1$, reducing profits in that period. Each unit of stock contributes $\frac{dF(x_{t+1})}{dx_{t+1}}$ units of growth. To offset the *direct* effect of the unit of increased harvest in period t, the owner must reduce harvest in period $t + 1$ by one unit; in addition, the owner must reduce harvest in period $t + 1$ by $\frac{dF(x_{t+1})}{dx_{t+1}}$ to make up for the reduced growth in period $t + 1$. The term $\frac{dF(x_{t+1})}{dx_{t+1}}$ corresponds to the interest payment in the savings example.

Each unit of reduced harvest in period $t+1$ reduces profits by $\left(p_{t+1} - \frac{C}{x_{t+1}}\right)$. Therefore, the reduction in period-$t + 1$ profits, caused by the *reduced* $t + 1$ harvest, is

$$\left(p_{t+1} - \frac{C}{x_{t+1}}\right)\left(1 + \frac{dF\left(x_{t+1}\right)}{dx_{t+1}}\right),$$

which equals the underlined term on the right side of equation 15.4. The lower $t+1$ stock (caused by the perturbation) increases harvest cost by

$$-\frac{\partial\left(\frac{C}{x_{t+1}}y_{t+1}\right)}{\partial x_{t+1}} = \frac{C}{x_{t+1}^2}y_{t+1},$$

the under-bracketed term. The time t present value cost of the perturbation equals the right side of equation 15.4.

15.1.2 Rent

We use the definition of rent to write the Euler equation more concisely. As in the nonrenewable resource setting, we define rent as the difference between price and marginal cost. For our example (but not in general), marginal cost equals average cost, so rent equals profit per unit of harvest.

For stock-independent average costs, rent is

$$R_t = p_t - C.$$

Using this definition, we rewrite the Euler equation 15.3 as

$$R_t = \rho R_{t+1}\left(1 + \frac{dF\left(x_{t+1}\right)}{dx_{t+1}}\right). \tag{15.5}$$

For stock-dependent harvest costs, rent is

$$R_t = p_t - \frac{C}{x_t} \tag{15.6}$$

and the Euler equation 15.4 becomes

$$R_t = \rho\left[\underline{R_{t+1}\left(1 + \frac{dF\left(x_{t+1}\right)}{dx_{t+1}}\right)} + \underbrace{\frac{C}{x_{t+1}^2}y_{t+1}}\right]. \tag{15.7}$$

The sole owner never sells where price is below marginal cost, so rent is never negative, and often (but not always) is positive. The open access fishery eliminates profits, driving rent to zero. Chapter 12.3.2 notes that Individual Transferable Quotas (ITQs) create property rights, making an open access or common property fishery more like a sole owner fishery. The equilibrium annual lease price of an ITQ equals that amount of profit a fisher can expect to obtain for the volume of harvest covered by the quota licence. The lease price thus provides an estimate of the rent associated with that volume of fish. Most lease prices range from 50% to 80% of the ex vessel fish price (the price fishers receive). Rent accounts for a substantial fraction of the value of fisheries protected by ITQs.

15.2 Policy

Objectives and skills:

- Understand why the optimal stock-dependent tax under *open access* equals the rent for the agent who harvests at the *first best* level.

Chapter 9.2 explains how to find an optimal tax in the presence of a market failure (e.g. market power or pollution): we first find the socially optimal level of output, and we then find a tax that "supports" this level of output. We say a tax "supports outcome X" if the market equilibrium in the presence of the tax is the same as "outcome X". The construction of optimal taxes in the dynamic setting follows the same logic. The First Fundamental Welfare Theorem (Chapter 2.6) states that, absent market failures, the price-taking sole owner harvests efficiently. Therefore, by solving the sole owner's problem we obtain the socially optimal trajectory. Information about that trajectory enables us to find the tax that supports the optimal trajectory under a particular market failure, such as open access.

We consider two scenarios. In the first, there is a single market failure: lack of property rights to the fishery. In order to find the tax that supports socially optimal harvest, we first find rent under the sole owner. We then note that if open access fishers are charged a tax, per unit of catch, equal to this level of rent, open access fishers harvest at the optimal level.

In the second scenario, the biomass provides ecological services that fishers ignore. For example, the stock being harvested might be important (perhaps as a food source) to another valuable fish stock. An owner who does not receive compensation for these services treats them as external to her decision problem, leading to excessive harvest and driving the stock to too low a level. This outcome leads to under-provision of the ecological services, creating a role for regulation even under the sole owner. Under open access, the optimal tax must correct the two market failures: the absence of property rights and the ecological externality.

We describe this policy problem in the fishery context, but the goal is for readers to become accustomed to thinking about common features of resource problems. Forests and fisheries give rise to similar market failures. There may be imperfect property rights in both cases; regardless of property rights to the physical resource, there may be unpriced benefits that are external to harvesters. Forests contribute to biodiversity and they sequester carbon. Absent policy intervention, foresters do not obtain these benefits. What is the right policy response when these types of externalities are important? How does that policy response depend on the nature of property rights? This section helps readers develop the skills needed to think systematically about these kinds of questions.

15.2.1 Optimal policy under a single market failure

Here we assume that there are no externalities, so the sole owner harvests efficiently. Under open access, there is a single market failure, arising from the absence of property rights to the fish stock. Our goal is to find a stock-dependent unit tax, a function $\nu(x)$, that induces open-access fishers to harvest efficiently. If open access fishers face the tax $\nu(x)$, they harvest up to the point where their tax-inclusive profits equal zero:

$$p(y_t) - \frac{C}{x_t} - \nu(x_t) = 0. \tag{15.8}$$

Comparing equations 15.6 and 15.8, we see that the levels of the price and harvest are the same in the two cases if and only if the open-access tax equals the sole owner's rent:

$$\nu(x_t) = R_t. \tag{15.9}$$

Unfortunately, in the open-access fishery we do not observe the sole owner rent, so we have to estimate it. If we can estimate the demand and harvest cost function, the growth function, and the initial stock, we can (numerically) solve the sole owner problem, and calculate the stock-contingent optimal tax, equal to the sole owner's rent, $R(x)$.

15.2.2 Optimal policy under two market failures

In this scenario, the stock of fish provides ecological services with per-period value $V(x_t)$, external to the sole owner. Because of this externality, harvest under the sole owner is not efficient (socially optimal). The efficient level of harvest maximizes

$$\sum_{t=0}^{\infty} \rho^t \left[\left(p_t - \frac{C}{x_t} \right) y_t + V(x_t) \right]. \tag{15.10}$$

The payoffs in equations 15.2 and 15.10 are identical, apart from the term $V(x_t)$ in the latter. Harvest under the sole owner is efficient if this owner receives a state-contingent subsidy $V(x_t)$.

The Euler equation when the sole owner receives the subsidy $V(x_t)$ is

$$R_t = \rho \left[\underline{R_{t+1} \left(1 + \frac{dF(x_{t+1})}{dx_{t+1}} \right)} + \underbrace{\frac{C}{x_{t+1}^2} y_{t+1}} + \underline{\underline{\frac{dV(x_{t+1})}{dx_{t+1}}}} \right]. \tag{15.11}$$

The right sides of equations 15.7 and 15.11 are identical, except for the double-underlined term in equation 15.11. The cost of a change, in period $t+1$, that offsets an additional unit of harvest at t, equals the present value of three terms. The single underlined term equals the loss in $t+1$ profit due to the reduced $t+1$ harvest; the under-bracketed term equals the higher cost due to the lower stock; the double-underlined term equals the reduced subsidy due to the lower stock.

Given estimates of the inverse demand function, the harvest cost function, and growth function, and the external benefit $(V(x))$ we can numerically solve the optimization problem and calculate rent in the presence of the subsidy $V(x)$. We denote this function as $\tilde{R}(x)$ instead of $R(x)$ to recognize that $V(x)$ alters the solution. Using the same reasoning as in Chapter 15.2.1, the stock contingent optimal tax for the open access fishery is $\tilde{R}(x)$.

In summary, the optimal tax for the open access fishery equals the rent (a function of the stock of fish) in the scenario *with no market failures.* If the only market failure under open access is the lack of property rights, we "merely" have to find the rent function, $R(x)$, under the sole owner. If there is an additional market failure, e.g. arising from ecological services that are external to the resource owner, then we have to find the rent function that would arise if the sole owner were induced to internalize the externality.

15.2.3 Empirical challenges

Managers are unable to directly observe the growth function, the cost function, or the biological stock. The management tools described above require estimates of these functions and the stock. We discuss two approaches to estimating these ingredients. The first relies on catch and effort data and simple models; the second approach uses more data and more complicated models.

To explain the first approach, consider the case where we have "panel data" of effort and catch for many boats for many years. The effort data consists of characteristics of the boats and expenditures; here, "effort" is a vector, not a single number. In a particular year, all of the boats confront (approximately) the same level of the stock. Using this fact and the panel data, we can estimate parameters relating harvest to the effort characteristics, and also estimate a "stock index".[1] This index involves both the physical stock and economic parameters; we cannot separate these, so we have a stock index, instead of an estimate of the actual stock.

The second approach is more complicated and requires more data, but permits the estimation of more complex models involving different species and different age or size categories within a species. Scientists collect samples by dragging nets across fishing areas. By counting rings in the bones (e.g. in the ear canal or the jaw), they estimate the age of individuals in the sample, in much the same way that counting rings of tree identifies the tree's age.

This data is used with a dynamic model and a measurement model. The dynamic model consists of a system of equations that describe the evolution of the stock(s) and age classes. Taking as given the equations' functional form, the goal is to estimate the parameters of the functions. In the simplest case, with a single stock variable, scientists might assume the logistic growth function, and then estimate the two parameters of that function, the natural growth rate and the carrying capacity. Actual applications tend to be much more complicated. The measurement model relates the underlying but unobservable variables of interest (e.g. stock size for different ages) to the measured variables (e.g. age estimates of the sample).

Estimation of the unknown parameters and unknown stocks uses an iterative procedure. Beginning with a guess of the unknown values of the parameters and the stocks, we can calculate what the measurements would have been, had the guess been correct. Of course, the guess is not correct, so the calculated values differ from the observed measurements. We then

[1] An "index" is a measurement related to the object of interest (here, the stock), not the object itself. For example, the economy-wide price level is a theoretical construction, not an observable price. We use the consumer price index to measure this price level, and then to measure inflation.

change our guess (using numerical methods) of the unknown values, in an effort to make the calculated values closer to the observed measurements. We stop the iteration when we decide that it is not possible to get a closer fit between the calculated values and the observed measures. The final iteration yields estimates of the parameters and stock variables.

15.3 The steady state

Objectives and skills:

- Obtain and analyze the steady state under the sole owner.
- Compare steady states under the sole owner and under open access.

Comparison of the sole owner and the open access steady states provides information about the relation between property rights and equilibrium outcomes. Under what circumstances do open access and sole ownership lead to the same (or almost the same) steady state? When are the steady states significantly different?[2]

By definition, at a steady state the harvest, stock, and rent are unchanging over time. We drop the time subscripts to indicate that these variables are constant in a steady state. The equation of motion of the stock is $x_{t+1} - x_t = F(x_t) - y_t$. In a steady state, the left side of this equation is zero; dropping the time subscripts, we write this equation, evaluated at the steady state, as $0 = F(x) - y$. We write the definition of the steady state rent as $R = p(y) - \frac{\partial c(x,y)}{\partial y}$. We obtain a third equation by evaluating the Euler equation at a steady state. We then have three algebraic equations in three unknowns, the steady state stock, harvest, and rent. We can solve these three equations to determine the steady state values. We consider the cases of constant and stock-dependent average extraction costs separately.

[2]In Chapter 14 we obtained the harvest rule under open access by solving the zero-profit condition, price = average cost. There, we could easily determine which of the steady states is stable. In the sole owner setting, we only have an optimality condition (the Euler equation), not an explicit harvest rule. We can still identify the steady states, but determining their stability requires methods discussed in Chapter 16.

15.3.1 Harvest costs independent of stock

We begin by evaluating equation 15.5 at a steady state (mechanically, dropping the time subscripts), to obtain

$$R = \rho R \left(1 + \frac{dF(x)}{dx}\right).$$

Using $\rho = \frac{1}{1+r}$, we multiply both sides of the equation by $1+r$ to obtain

$$(1+r)R = R\left(1 + \frac{dF(x)}{dx}\right) \Rightarrow rR = R\left(\frac{dF(x)}{dx}\right) \Rightarrow$$

$$0 = R\left(r - \frac{dF(x)}{dx}\right). \tag{15.12}$$

Equation 15.12 implies that at a steady state either $R = 0$ or

$$r = \frac{dF(x)}{dx}. \tag{15.13}$$

Equation 15.13 (a special case of the "modified golden rule") states that at an interior steady state with $R > 0$, the sole owner is indifferent between two investment opportunities. The owner can increase current harvest, and invest the additional profit in an asset that earns the annual return r, or she can keep the extra unit of stock in the fishery, where it contributes to growth, thus contributing to future harvests and future profits. At an interior optimum, the owner is indifferent between these two investments.

Provided that F is concave, there is a unique solution to equation 15.13, denoted x_∞, which depends only on the discount rate and the growth function. If $x_\infty < 0$, we conclude that there is no interior steady state with positive profits. If $x_\infty > 0$ we perform one further test. The harvest that maintains x_∞ as a steady state is $y_\infty = F(x_\infty)$. The price at this level of harvest (using the inverse demand function) is $p(y_\infty)$ and the rent is $R(y_\infty) = p(y_\infty) - C$. If $x_\infty > 0$ and $R(y_\infty) \geq 0$, then x_∞ is a steady state. If either of these inequalities fail, x_∞ has no significance, and there are no interior steady states with positive profits. There might still be interior steady states with zero profits.

In summary, the solution to equation 15.13, x_∞, is a steady state if and only if it satisfies both of the inequalities

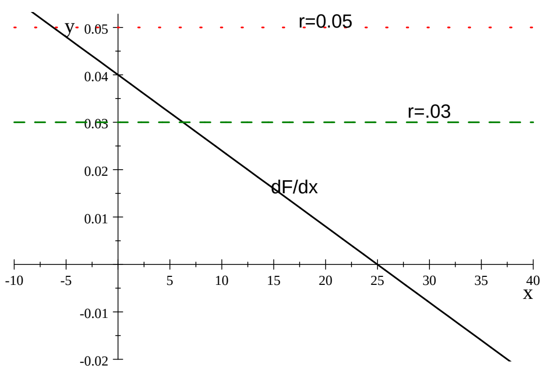

Figure 15.1: The derivative of the logistic growth $F(x) = 0.04x\left(1 - \frac{x}{50}\right)$, $\frac{dF}{dx} = 0.04x\left(1 - \frac{2x}{50}\right)$ and two values of r.

$$(i)\ x_\infty > 0, \text{and}\quad (ii)\ R(y_\infty) \geq 0. \tag{15.14}$$

We illustrate this procedure for determining interior steady states using the logistic growth function, $F(x) = \gamma x\left(1 - \frac{x}{K}\right)$ and Figures 15.1 and 15.2.

Figure 15.1 shows the graph of $\frac{dF(x)}{dx} = \gamma\left(1 - \frac{2x}{K}\right)$ for $\gamma = 0.04$ and $K = 50$. The intercept of this graph is $\gamma = 0.04$, the intrinsic growth rate. The figure also shows two horizontal lines labelled $r = 0.05$ and $r = 0.03$. The intersection of each of these lines and the graph of $\frac{dF}{dx}$ is the solution to equation 15.13 for the particular value of r. For $r = 0.05$, this solution occurs where $x_\infty < 0$. Because the stock cannot be negative, we conclude that for $r = 0.05$ (or any value $r \geq \gamma$) there is no interior steady state with positive profits. For $r = 0.03$ (and also for any $r < \gamma$), the solution to equation 15.13 occurs where $x_\infty > 0$. There is a positive solution ($x_\infty > 0$) to equation 15.13 if and only if $r < \gamma$. This inequality implies that the value of harvesting an additional fish is less than the value of allowing the fish to remain alive and reproduce.

To find x_∞, we use $\frac{dF(x)}{dx} = \gamma\left(1 - \frac{2x}{K}\right)$ and solve $r = \gamma\left(1 - \frac{2x}{K}\right)$ to obtain

$$x_\infty = \frac{K}{2}\left(1 - \frac{r}{\gamma}\right) < \frac{K}{2}. \tag{15.15}$$

As noted above $x_\infty > 0$ if and only if $r < \gamma$. The candidate steady state decreases with $\frac{r}{\gamma}$: a higher discount rate (greater impatience) lowers the candidate, and faster growth increases the candidate. The inequality in equation

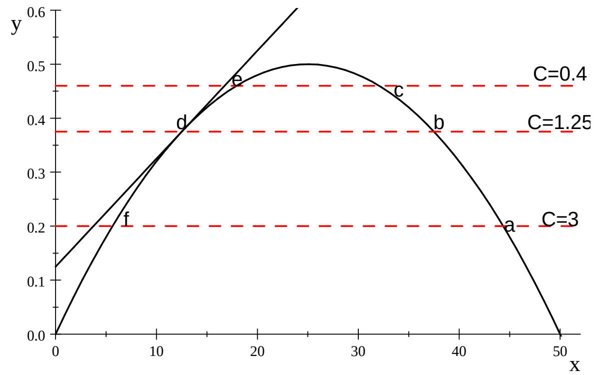

Figure 15.2: For $F = 0.04x\left(1 - \frac{x}{50}\right)$ and $r = 0.02$ one candidate for a steady state under the sole owner is point d, where $x = 12.5$ and $y = .375$. For inverse demand $= 5 - 10y$, open access steady states (where rent and growth are both zero) are points e and c for $C = 0.4$, points d and b for $C = 1.25$, and points f and a for $C = 3$. These points are also steady states under the sole owner for these levels of C.

15.15 states that the candidate steady state is less than $\frac{K}{2}$, the stock level that maximizes growth (leads to Maximum Sustainable Yield, MSY). Thus, at a steady state with positive rent, harvest is less than MSY.

A numerical example We use the logistic growth function $F(x) = 0.04x\left(1 - \frac{x}{50}\right)$ and $r = 0.02$. Here, $\gamma = 0.04 > r$, so $x_\infty > 0$. We find x_∞ by solving equation 15.15 to obtain $x_\infty = 12.5$ and then obtain

$$y_\infty = F(x_\infty) = 0.04\,(12.5)\left(1 - \frac{12.5}{50}\right) = 0.375.$$

Figure 15.2 identifies the candidate steady state as point d, $(x_\infty, y_\infty) = (12.5, 0.375)$, the tangency between the graph of the growth function and the line with slope $r = 0.02$. To determine whether the candidate is a steady state, we check whether rent is positive at this value, i.e. whether the candidate satisfies equation 15.14, part (ii). Here we use the inverse demand function $p(y) = 5 - 10y$. With the choke price 5, the fish has value if and only if $C < 5$, as we hereafter assume. Rent at the candidate steady state, point d, is $R = 5 - 10(.375) - C$: $R \geq 0$, requires $C \leq 1.25$.

Figure 15.2 shows three horizontal lines, each of which gives the level of harvest at which rent is 0, for a particular value of C. Rent is zero if $R = 5 - 10y - C = 0$, implying $y = \frac{5-C}{10}$; for example, the intercept of the horizontal line labelled $C = 0.4$ is $\frac{5-0.4}{10} = 0.46$. For a given value of C, *rent is positive below the* $R = 0$ *line, and rent is negative above this line.*

We now have the information needed to determine whether $R(y_\infty) \geq 0$. Point d, our candidate steady state, lies below the $R = 0$ line for $C = 0.4$. Thus, if $C = 0.4$ (harvest costs are low), rent is positive at point d; in this case, point d is a steady state. In contrast, if $C = 3$ (harvest costs are high), rent is negative at point d; for this level of costs, point d has no significance. If $C = 1.25$, rent is 0 at point d; for these costs, point d is an interior steady state at which rent is 0.

We also use Figure 15.2 to discuss the relation between the open access and the sole owner fisheries. From Chapter 14.1.1 we know that under open access there are three steady states: 0, an intermediate steady state (shown as the values e, d, f for the three values of C) and a high steady state (c, b, a for the three cases). Under open access the high steady state and $x = 0$ are stable steady states, and the intermediate steady state is unstable.

Now consider the steady states under the sole owner. The open access high steady states, (c, b, a) also satisfy the sole owner steady state conditions, for the different levels of cost, C. At these points: (i) $y = F(x)$, so the stock is unchanging, (ii) $R = 0$, so the steady state Euler equation 15.12 is satisfied, and (iii) the definition of rent is also satisfied. Exactly the same reasoning holds at the open access intermediate steady states, and at $x = 0$. Thus, for this problem, all of the steady states under open access are also steady states under the sole owner. If $C < 1.25$, then point d is also a steady state under the sole owner (but not under open access).

In summary, every point that is a steady state under open access is also a steady state under the sole owner. (This conclusion holds for constant average harvest costs, but not for general harvest costs.) At all of these points, either rent or harvest is zero. For sufficiently low costs ($C < 1.25$ in our example) there is an additional steady state under the sole owner, that does not exist under open access. At that steady state, the sole owner has positive rent. The next chapter turns to the question of determining which of the sole owner steady states is stable.

A second numerical example Suppose that $F = 10x\left(1 - \frac{x}{100}\right)$, $r = 8$, inverse demand is $p = 60 - y$ and the constant extraction cost is C. To obtain x_∞, set the derivative $\frac{dF}{dx} = 10\left(1 - \frac{x}{50}\right)$ equal to $r = 8$ and solve for x: $10\left(1 - \frac{2x}{50}\right) = 8 \Rightarrow x_\infty = (10 - 8)\frac{5}{2} = 5$. For this example, the candidate steady state is positive, i.e. it passes the first of the two tests in equation 15.14. We find harvest at the steady state by solving $y_\infty = F(x_\infty) = 10(5)\left(1 - \frac{5}{100}\right) = 47.5$. Next we check whether rent is non-negative at this candidate. Rent is $R_\infty = 60 - 47.5 - C = 12.5 - C$. We conclude that rent is non-negative (so $(x_\infty, y_\infty) = (5, 47.5)$ is a steady state) if and only if $C \leq 12.5$. If $C > 12.5$ the stock $x = 12.5$ has no significance.

Box 15.1 Sensitivity of the steady state to the discount rate The steady state stock with positive rent tends to be more sensitive to the discount rate, the more slowly the stock grows. Using equation 15.15, the elasticity of the steady state stock, with respect to the discount rate in the logistic model, is $\frac{r}{\gamma - r}$. For fast-growing Pacific halibut, γ is estimated (with an annual time step and a continuous time model) at 0.71; for the slow-growing Antarctic Fin-whale the estimate is 0.08. As r ranges from 2%-5%, the elasticity of the steady state for the Fin-whale is 11–22 times greater than the elasticity for halibut: the steady state of the more slowly growing stock is more sensitive to the discount rate.

15.3.2 Harvest costs depend on the stock

When harvest costs depend on the stock we show that:

- For low harvest costs, the sole owner steady state stock is lower than the stock that maximizes steady state yield; large harvest costs reverse this relation.

- Higher harvest costs might either increase or decrease the owner's steady state rent, or the steady state consumer surplus.

We use equation 15.7 to write the steady state condition for rent (merely by dropping the time subscripts):

$$R = \rho\left[R\left(1 + \frac{dF(x)}{dx}\right) + \frac{C}{x^2}y\right].$$

Using $\rho = \frac{1}{1+r}$, multiplying both sides by $1+r$ and cancelling terms, we simplify this equation to obtain

$$\left(r - \frac{dF(x)}{dx}\right) R = \frac{C}{x^2} y.$$

We collect the steady state conditions in

$$F(x) - y = 0, \quad \left(r - \frac{dF(x)}{dx}\right) R - \frac{C}{x^2} y = 0, \text{ and } R = p(y) - \frac{C}{x}. \tag{15.16}$$

The first two equations repeat the steady state conditions for the stock and the rent, and the third equation repeats the definition of rent.

System 15.16 comprises three equations in three unknowns, x, y, and R. All three equations must hold if $x > 0$, i.e. at an interior solution; in addition, $R \geq 0$ must be satisfied.[3] Figure 15.3 shows the graph of the growth function and the graphs of the steady state Euler equation (the upward sloping curves) corresponding to different values of C. (The figure uses $p = 10 - y$, $F = 0.04x\left(1 - \frac{x}{50}\right)$ and $r = 0.02$.) At all points on the growth function, harvest equals growth: the first equation in system 15.16 is satisfied. At all points on an upward sloping curve (corresponding to a particular value of C), the second two equations in the system are satisfied. Thus, satisfaction of all three equations occurs (only) at a point of intersection. That intersection is the sole owner steady state (with positive profits).

The figure shows that an increase in C causes the graph of the steady state Euler equation to shift to the right, increasing the value of x_∞: larger values of the cost parameter lead to larger steady state fish stocks. Figure 15.3 also shows that the sole owner steady state might lie either to the left or the right of the MSY stock level. Discounting gives the sole owner (and the social planner) an incentive to harvest earlier rather than later, tending to decrease the steady state stock. The responsiveness of harvest cost to the stock gives the owner (and the planner) an incentive to build up the stock in order to decrease future harvest cost, tending to increase the steady state stock. The discounting incentive dominates for small C and the cost incentive dominates for large C.

---

[3]For $C > 0$, the second equality in system 15.16 implies $R = \left(\frac{C}{x^2} y\right) / \left(r - \frac{dF(x)}{dx}\right) > 0$, if and only if $\left(r - \frac{dF(x)}{dx}\right) > 0$. For $C = 0$, there are two candidate steady state, one of which satisfies $r - \frac{dF(x)}{dx} = 0$; we have to confirm that this candidate is actually an equilibrium, by checking that $p(y_\infty) \geq 0$.

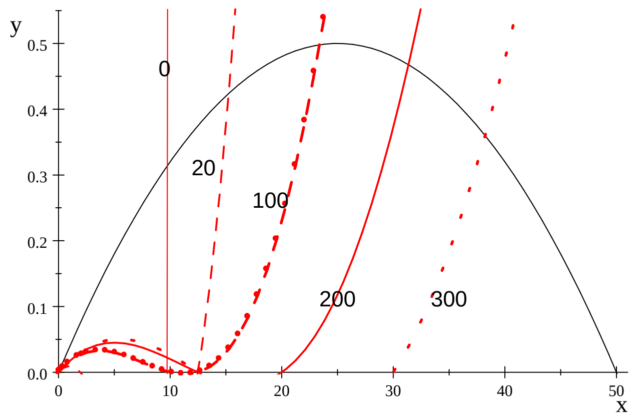

Figure 15.3: Graphs of F and the steady state condition for rent, the upward sloping curves (each labelled with the value of the cost parameter, C). The sole owner steady state occurs at the intersection of the curves, (x_∞, y_∞). An increase in C increases x_∞ an has an ambiguous effect on y_∞.

C	x_∞^{so}	y_∞^{so}	R_∞^{so}	x_∞^{oa}	y_∞^{oa}
0	12.5	0.38	9.6	0	0
50	18.6	0.47	1.7	5.1	0.18
300	38.2	0.36	1.8	31.47	0.47

Table 15.1 Steady state stock, harvest, and rent for the sole owner, and stock and harvest under open access.

As C increases, moving (x_∞, y_∞) to the right, the harvest level, y_∞, rises and then falls: there is a non-monotonic relation between C and the steady state harvest. Table 15.1 shows sole owner (superscript "*so*") steady state values of the stock, harvest, and rent, under three values of the cost parameter, C. The last two columns of the table show the open access (superscript "*oa*") steady state stock and harvest for those values of C. Open access rent is always zero. Steady state harvest, and thus consumer surplus, might be either higher or lower under the sole owner, compared to open access.

The price - harvest combination always lies on the demand function. Higher harvests therefore correspond to lower prices and higher consumer welfare. Because sole owner steady state harvest is non-monotonic in C, a higher C might make consumers either better or worse off in the steady state. In the static competitive setting, in contrast, higher costs shift the equilibrium supply function in and up, leading to a higher equilibrium price and lower consumer surplus. Why do higher costs have different effects in the (steady state) sole owner fishery, compared to the static market? The answer rests on two facts. First, in the fishery context, a higher C reduces the incentive to harvest, leading to a higher steady state stock. Second, the higher steady state stock might correspond to either a higher or a lower steady state harvest, depending on whether the stock lies below or above the level corresponding to MSY.

Steady state rent $(p(y) - \frac{C}{x})$ is also non-monotonic in the cost parameter. For a given stock, a higher value of C increases costs, $\frac{C}{x}$, lowering rent. However, the steady state values of y and x both change with C. If the steady state lies to the left of the MSY level, higher C increases harvest, reducing the price, thus reducing rent; to the right of the MSY level, the higher C reduces harvest, thus increasing the price. Table 15.1 shows that steady state rent is lower at $C = 50$ compared to either $C = 0$ or $C = 300$.

15.3.3 Empirical evidence

Is the optimal steady state stock above or below the MSY stock? It is not surprising that the empirical answers are inconclusive, because they depend on extraction costs, the growth function, and the demand function; these differ across fisheries, and are difficult to measure. The answer also depends on the discount rate, r, about which there is disagreement.

A 2007 study of four fisheries, including slow-growing long-lived orange roughy, finds that the socially optimal stock level *exceeds* the MSY stock level. Therefore, the steady state harvest is below the MSY. The fact that the study includes a slow-growing fish is important. Chapter 15.3.1 notes that a low growth rate reduces the optimal steady state stock, and possibly leads to extinction. Because other considerations (e.g. strongly stock-dependent harvest costs) cause the steady state to increase, the optimal steady state depends on a balance of conflicting forces. A 2013 study for North Pacific albacore, concludes that the optimal steady state lies *below* the MSY level. This study emphasizes the role of cost-reducing technology improvements. As Figure 15.3 illustrates, lower harvest costs (smaller C) reduce the sole owner steady state.

Many actual management practices try to keep the stock at the level of MSY. There is a plausible and a dubious argument in favor of this practice. The plausible argument is that, lacking a strong *a priori* basis for thinking

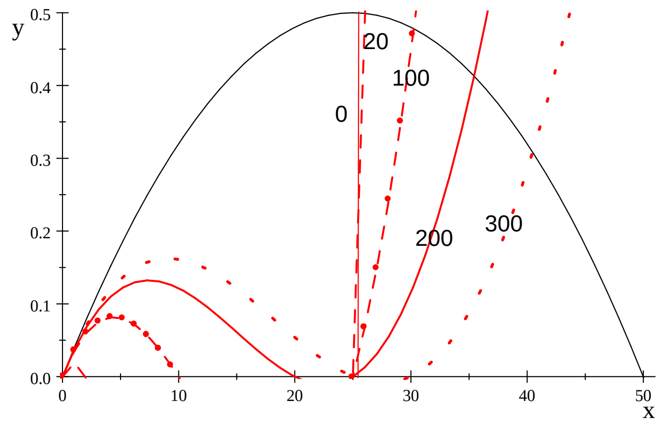

Figure 15.4: Graphs of F and the steady state Euler equations. Parameter values as in Figure 15.3 except that here $r = 0$.

that the stock should be to the right or the left of this level, and in view of the measurement difficulties, the MSY level is "neutral". The dubious argument, based on intergenerational ethics, is that a positive discount rate is unfair to future generations, because it gives them less weight in the social welfare function. Even it one accepts this view of ethics, it does not imply that the ethically optimal steady state occurs at MSY. That level maximizes consumer surplus, but when harvest cost depends on the stock it does not maximize social welfare, the sum of consumer and producer surplus. With stock dependent costs, low discount rates require a higher steady state stock to take advantage of cost reductions. Because these stocks occur to the right of the MSY level, they correspond to lower harvests and lower consumer surplus, but higher producer surplus, and a higher social surplus.

Figure 15.4 illustrates this claim, reproducing the graphs in Figure 15.3, but replacing $r = 0.02$ (a 2% per annum discount rate) with $r = 0$. For stock-independent costs ($C = 0$), the optimal steady state (for $r = 0$) occurs at the MSY. However, if harvest costs depend on the stock, the optimal steady state stock always lies to right of the MSY. For any value of C, a decrease in the discount rate increases the optimal steady state stock.

15.4 Summary

We derived and interpreted the Euler equation for the price-taking sole-owner fishery, emphasizing the difference between the renewable and nonrenewable resources. For a renewable resource, we have to take growth into account. The sole owner internalizes the effect of her current harvest decisions on future stocks. Absent externalities, the First Fundamental Welfare Theorem implies that the outcome under the sole owner is efficient. In that case, there is no efficiency rationale for regulation. Moving from open access to the sole owner solves the only market failure.

If it is not possible or politically desirable to privatize an open access fishery (thus moving to the sole-owner scenario), the open access fishery can be induced to harvest efficiently by charging a tax per unit of harvest equal to rent under the sole owner. We also considered the case where the stock of fish provides ecological services that are external to the sole owner. A subsidy or a tax can induce the sole owner to internalize that externality, in which case the sole owner again harvests efficiently. A tax equal to the sole owner's rent, when that owner harvests efficiently, induces the *open access* fishery to harvest efficiently. Because this efficiency-inducing tax varies with the stock of fish, a constant tax is not first best.

For the nonrenewable resource, extraction eventually ceases as the resource is exhausted or as extraction becomes too costly to be economically rational. For the renewable resource, the sole owner might drive the stock to a positive steady state, where harvest and the stock remain constant forever; or the owner might drive the fishery to extinction. A model with constant (stock-independent) average harvest costs illustrates these possibilities. Here, if the *intrinsic* growth rate is less than the rate of interest ($\gamma < r$), there is no interior steady state with positive rent: either the owner drives the stock to extinction, or she maintains the stock at a positive level with zero rent. If the intrinsic growth rate exceeds the rate of interest ($\gamma > r$), there is a candidate interior steady state at which the *actual* growth rate equals the rate of interest. This candidate is a steady state for the sole owner if and only if rent is greater than or equal to zero there. In this case, the owner does not drive the stock to extinction. There is typically another interior steady state at which rent is zero.

Stock dependent harvest costs give the sole owner an incentive to restrict harvest in order to let the stock grow, thus reducing future harvest costs. Thus, stock dependence tends to increase the sole owner steady state, while discounting tends to decrease it. The sole owner steady state might lie above or below the MSY stock level. Lower discount rates increase the sole owner steady state. In the limiting case with zero discounting, the sole owner steady state equals the MSY level when harvest costs do not depend on stocks; with stock dependent harvest costs, that steady state is above the MSY stock level.

15.5 Terms, study questions, and exercises

Terms and concepts

stock-dependent efficiency-inducing tax, ecological services

Study questions

1. Given a growth function $F(x)$, a cost function $c(x, y)$ and discount factor ρ write down the sole owner's objective function and constraints. Without doing calculations, describe the steps needed to obtain the Euler equation in this model.

2. If you are given the Euler equation for a particular model (with or without stock dependent harvest costs) you should be prepared to provide an economic interpretation of this equation.

3. (a) Consider the case where marginal harvest costs equal average harvest costs. Identify the stock-dependent tax that induces the open access industry to harvest at the same rate as the untaxed sole owner. (Compare the equilibrium conditions under the sole owner (price – marginal costs $=$ rent) and under open access (price - average costs $= 0$). (b) Suppose instead that harvest costs are convex in harvest (so that marginal costs exceed average costs). In order to induce the open access fishery to harvest at the same level as the untaxed sole owner, would you have to increase or decrease the tax you identified in part (a)? Explain.

4. For the model with constant (stock independent) average harvest costs, C, and logistic growth $F(x) = \gamma x \left(1 - \frac{x}{K}\right)$, the Euler equation evaluated at the steady state is $0 = R\left(r - \frac{dF(x)}{dx}\right)$. (a) What is the

definition of steady state rent? (b) Under what conditions is there an interior steady state with positive rent? Explain. (c) If the conditions in part (b) are not satisfied, are there other interior steady states? If so, describe these.

5. Consider the following two claims "(i) An open access fishery leads to excessive harvest. (ii) It is never socially optimal to exhaust the stock." Discuss these two claims in light of the model described in Question 4. State whether you agree or disagree with these claims and justify your answer.

Exercises

1. (a) Derive the Euler equation for the sole owner price-taking fishery when average costs are constant, independent of the stock. (b) Provide an intuitive explanation for this equation. (Students can answer this question by mimicking the derivation and the explanation in the text, making changes to reflect the different cost function here.)

2. Write down and interpret the Euler equation for the monopoly owner of a renewable resource facing inverse demand $p(y)$ and constant average harvest costs, Cy. (Hint: use the same methods that we applied to the nonrenewable resource problem.)

3. Suppose that the sole owner receives a subsidy equal to $V(x_t)$ in each period. the stock provides ecological services $V(x_t)$, and ; that subsidy internalizes the externality. Provide the economic intuition for the Euler equation 15.11. You need to understand the intuition for the simpler case (where V is absent) and then explain why the presence of payment for the ecological services changes the optimality condition.

4. Derive the formula for the elasticity of the steady state with respect to the discount rate presented in Box 15.1.

5. Show that, provided that $F(x)$ is concave, an interior steady state in the case where costs do not depend on the stock, always lies below the maximum sustainable yield. A price-taking sole owner has the discount rate $r = 0.03$, implying the discount factor $\rho = \frac{1}{1+0.03} \approx 0.97$. (a) Write down the sole owner's optimization problem. (b) Write the Euler equation, using the inverse demand function to replace p_t, the

growth function, and $\rho = 0.97$. (c) Write down the Euler equation and the growth function evaluated at the steady state. (d) Write the definition of steady state rent, and use this definition to write the steady state Euler equation. (e) Follow the steps that produce equation 16.12, using the functional forms and parameter values given above. (f) Now set $C = 4.5$ and identify all interior steady states.

6. Consider the model with constant harvest costs C and logistic growth $F(x) = \gamma x \left(1 - \frac{x}{K}\right)$. Suppose that $\gamma < r$. Are there circumstances where the sole owner drives the stock to a positive steady state? Explain and justify your answer. (Hint: Is there any reason to suppose that rent is positive in this model?)

7. Using equation 15.15 for the case of stock-independent harvest costs with logistic growth, verify that if $\gamma > r$, then a larger value of K or γ or a smaller value of r, all increase the interior steady state. Provide an economic (not mathematical) explanation for these results. You need to explain how changes in these parameters changes the owner's incentive to conserve the fish. Think about how an increase in r changes the current valuation of future rents. Think about how increases in K or γ alter the value of a larger fish stock.

Sources

Clark (1996) provides the estimates of growth rates used in Box 15.1.

Homans and Wilen (2005) provide the estimate of the annual lease prices for fishing quotas, as a percent of ex vessel price of catch.

Fenechel and Abbott (2014) show how estimates of stock dynamics and the production function can be used to estimate the gain from better management of fish stocks.

Zhang and Smith (2011) describe and implement, for Gulf Coast reef fish, the first estimation approach discussed in Chapter 15.2.3.

http://www.nmfs.noaa.gov explains the second estimation approach in Chapter 15.2.3.

The International Scientific Committee for Tuna (2011) illustrates the second estimation approach, for the case of tuna stocks.

Grafton et al. (2007) provide evidence that the socially optimal stock exceeds the MSY stock.

Squires and Vestergaard (2013) provide evidence that increases in technical efficiency cause socially optimal steady state stocks to be below the MSY stock.

Chapter 16

Dynamic analysis

Objectives

- Use the sole owner optimality conditions to study resource dynamics and resource policy.

Skills

- Use intuition and economic reasoning to analyze the case of constant average harvest costs.
- Use phase portrait analysis to analyze dynamics under stock dependent harvest cost.
- Compare harvest rules under the sole owner and under open access to characterize the efficiency-inducing tax for open access.

This chapter moves beyond steady state analysis to study the evolution of the fish stock and harvest under the sole owner. We assume throughout the chapter that the fishery provides no non-market (e.g. ecosystem) services, so there are no market failures under the price taking sole owner. The sole owner and the social planner make the same decisions. We compare the optimally controlled stock with the stock trajectory under open access in order to characterize the optimal open access tax.

The market failure under open access arises from the lack of property rights. As Chapter 12.3.2 emphasizes, the first best remedy in this situation (usually) involves institutional changes, e.g. the establishment of property rights. However, only a small fraction of fisheries are currently managed using property rights-based regulation. It is worth understanding how other types of regulation, such as taxes, can increase efficiency under open access.

We emphasize graphical analysis, using a parametric model. This qualitative analysis provides information about the direction of change and the relation between initial conditions and the ultimate steady state. Chapter 13.3.1 notes that the discrete time dynamics are complicated because the stock can "jump" from one side to the other of a steady state. The continuous time limit of the discrete state model is much easier to study, so we replace the (discrete time) difference equations with (continuous time) differential equations.

We consider a first scenario in which average harvest costs are constant, and a second in which harvest costs depend on the stock. The first scenario makes it possible to obtain results using economic reasoning, without introducing additional mathematical tools. This approach is useful for intuition, but it disguises many subtleties, and it does not suggest a method for analyzing more general problems. The case of stock-dependent costs requires additional tools, which are useful for a wide variety of dynamic problems.

It is important to keep in mind that the solution to the sole owner's problem is a "harvest rule", giving optimal harvest as a function of the stock; we represent a harvest rule by showing its graph. Chapter 13 uses exogenous harvest rules, ones without any basis in theory. Chapter 14 derives the endogenous harvest rule under open access, by finding the harvest, a function of the fish stock, that drives rent to zero. We obtained those rules merely by solving an equation (rent $= 0$). We denote the endogenous harvest rule for the sole owner's optimization problem as $y(x)$ (harvest as a function of the stock). Absent a closed form expression for this function, we rely on either qualitative or numerical analysis.

16.1 The continuous time limit

Objectives and skills

- Provide an intuitive understanding of the continuous time analog of the discrete time Euler equation.

We need one intermediate result: the continuous time version of the discrete time Euler equation. In deriving that equation, we did not specify whether the length of a period is one year or one second. Here we assume that the length of period in the discrete time setting is sufficiently small that the continuous time limit provides a reasonable approximation.

As in Chapter 13.3, the continuous limit of the discrete time equation of motion for the stock is $\frac{dx}{dt} = F(x) - y$. The missing piece is the continuous time Euler equation. We begin with the discrete time equation 15.5 for constant average harvest cost and equation 15.7 for stock dependent cost, and take limits, letting the length each period become small. The continuous time limits are, respectively[1] (Appendix J)

$$\frac{dR_t}{dt} = R_t\left(r - \frac{dF(x_t)}{dx_t}\right), \tag{16.1}$$

$$\frac{dR_t}{dt} = R_t\left(r - \frac{dF(x_t)}{dx_t}\right) - \frac{C}{x_t^2}y_t. \tag{16.2}$$

16.2 Harvest rules for stock-independent costs

Objectives and skills

- Characterize the sole owner harvest rule under constant costs.
- Compare this rule and the open access harvest rule to describe the optimal open access tax.

This section uses the example introduced in Chapter 15.3.1, with growth function $F(x) = 0.04x\left(1 - \frac{x}{50}\right)$, inverse demand $p(y) = 5 - 10y$, discount rate $r = 0.02$, and constant average costs C, the only free parameter. By varying C, we determine the relation between harvest costs and the sole-owner equilibrium. The Euler equation 16.1 must hold at every point along the sole-owner's harvest path.

[1] These continuous time Euler equations can be obtained using the calculus of variations, the Maximum Principle (employing Hamiltonians) or continuous time dynamic programming. These methods require additional mathematics. The approach here is heuristic, using only the discrete time perturbation method, and then taking a "formal limit", but without proving that this limit is mathematically valid. It does, however, give the correct optimality condition.

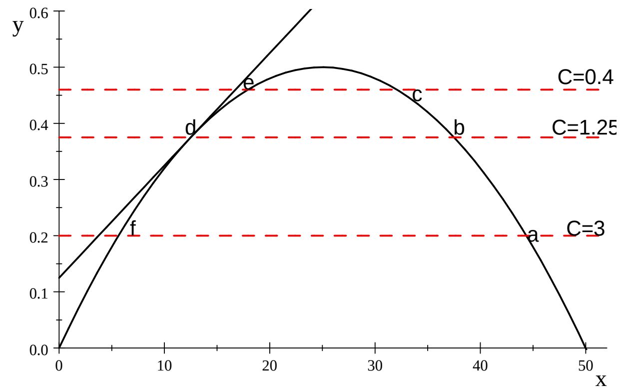

Figure 16.1: Open access and sole owner steady states for $F = 0.04x\left(1 - \frac{x}{50}\right)$, $r = 0.02$, and $p = 5 - 10y$.

We reproduce Figure 15.2, here shown as Figure 16.1. First consider the open-access equilibrium. The horizontal dashed lines in Figure 16.1 are the open-access harvest rules corresponding to three values of C. (These dashed lines show the open access harvest rules for $x > 0$; at $x = 0$, harvest is always zero. You cannot get blood out of a turnip.) These (constant) values of y satisfy the zero-profit open access condition, $5 - 10y = C$, or $y = \frac{5-C}{10}$. Using the analysis in Chapter 14, the points a, b and c are stable steady states, and f, d and e are unstable steady states, for the three values of C. The origin, $x = 0$ is a stable steady state in all three cases. For example, at $C = 0.4$, the open-access stock approaches point c if the initial stock is greater than the horizontal coordinate of point e; if the initial stock lies below this level, the open-access stock approaches $x = 0$. The entries in Table 1 show the open access steady state, x_∞, corresponding to different values of C and different initial conditions, x_0. In writing $x_\infty = a$, for example, we mean that x_∞ equals the *horizontal coordinate* of point a.

	x_0 above unstable steady state (e, d or f)	x_0 below unstable steady state (e, d or f)
$C = 0.4$	$x_\infty = c$	$x_\infty = 0$
$C = 1.25$	$x_\infty = b$	$x_\infty = 0$
$C = 3$	$x_\infty = a$	$x_\infty = 0$

Table 16.1: Open-access steady state, x_∞, for different initial conditions, x_0

We now consider the sole owner. Here we cannot find an explicit function for the harvest rule for *all* values of x; however we know that rent is nonnegative whenever extraction is positive: $y > 0 \Rightarrow R \geq 0$. This fact, and some reasoning discussed below, enable us to identify the steady state that the stock approaches, as a function of the value of C and of the initial conditions x_0. Table 2 summarizes this information, and Section 16.2.2 explains how we obtain it. First, we consider the policy implications of Tables 1 and 2.[2]

	x_0 above middle steady state (e, d or f)	x_0 below middle steady state (e, d or f)
$C = 0.4$	$x_\infty = c$	$x_\infty = d$
$C = 1.25$	$x_\infty = b$	$x_\infty = d$
$C = 3$	$x_\infty = a$	$x_\infty = f$

Table 16.2: Sole owner steady state, x_∞, for different initial conditions, x_0.

16.2.1 Tax policy implications of Tables 1 and 2

Tables 1 and 2 imply a simple and intuitive policy recommendation:

> It is important to regulate an open access fishery when the stock is small, but regulation may not be needed when the stock is large.

We say that the initial stock is "large" if it exceeds the middle steady state, points e, d or f (depending on the value of C); the initial stock is "small" if it is below these levels. Thus, the precise meaning of "large" and "small" depends on C. For all three values of C, if the initial stock is "large", the equilibrium is the same under open access and under the sole owner. In these cases, the stock is large enough that the sole owner's rent, along the equilibrium trajectory, is 0, exactly as under open access. Here, the resource is not scarce; there is no reason to tax open access harvest, because there is no market failure in this circumstance.

In contrast, if the initial stock is "small" (but positive), open access drives to the stock to extinction, whereas the sole owner drives the stock to the

[2]For $C = 1.25$ and $C = 3$, the sole owner middle steady states (d and f, respectively) are "semi-stable": for initial conditions to the left of these points the trajectory converges to the point (d or f, depending on the value of C), and for initial conditions to the right, the trajectory moves away from that point.

socially optimal positive level. Here, the resource is scarce; there is a market failure, and a need for regulation under open access. For $C < 1.25$, and small x_0, the sole owner drives the stock to point d. Rent is positive en route to, and at the steady state.[3] The open access drives the stock to extinction. Here, it is important to tax (or otherwise regulate) open access harvest in order to preserve the resource stock. The first best tax policy, for small stocks, varies with the level of the stock. We previously noted (Chapter 15.2) that in a resource setting, the optimal tax is typically stock-dependent. Optimality might be too much to ask for, but the analysis suggests second-best alternatives. The manager can close down the open access fishery until the stock recovers to point d, and then maintain a constant tax that supports open-access harvest at point d. A less extreme alternative uses a high tax to permit recovery of low stocks, reducing the tax as the stock increases.

Summary For this model, there is no need to regulate an open access fishery with large stocks, but regulation is important when the stock is small. Taxes can alter the steady state to which stock converges under open access, and also alter the speed at which the stock changes. At low stocks, the optimal open access tax depends on the stock. If it is impractical to use the optimal tax, second best taxes can insure that the stock approaches the first best steady state, even if the approach does not occur at the optimal speed.

16.2.2 Confirming Table 2 (*)

We use Figure 16.2 to examine the case $C = 0.4$, leaving the other two cases as exercises. Figure 16.2 contains a vertical line at $x = G$, through the unstable steady state; the initial stock is large if $x_0 > G$ and small if $x_0 < G$.

For $x \geq G$, the sole owner can harvest at the rate that drives profits to zero (equal to the open access rate). Beginning with $x > G$, if the owner harvests at this rate, the stock is driven to the higher steady state, point c. Along that trajectory, $R = 0$, so the Euler equation 16.1 is satisfied. For this problem, the necessary conditions for optimality are also sufficient, so for $x > G$ the sole owner harvests at the open access level.

[3]If $5 > C > 1.25$, rent is negative at point d. If x_0 is small, the sole owner drives the stock to a steady state to the left of and below d (e.g., point f for $C = 3$). In this case, rent is positive en route to the steady state, but zero at the steady state. The open access fishery drives the stock to extinction.

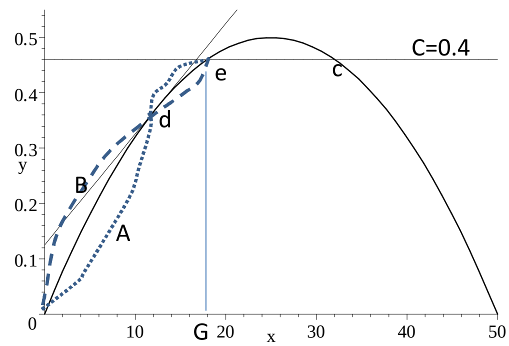

Figure 16.2: The growth function $F(x)$ and two hypothetical harvest rules (dashed and dotted) for $x < G$.

The interesting situation arises for $x < G$. We established (Chapter 15.3.1) that there is a unique candidate steady state with positive rent, point d. This point lies below the zero-profit (horizontal) line, so d satisfies both of the conditions in equation 15.14; it is the unique steady state with positive profit. The harvest rule must intersect the growth function at point d, but it cannot intersect the growth function at any other points below G. If there was such a point of intersection, that point would be a steady state with positive rent; but we know that d is the only such point.

Therefore, to the left of point d the harvest rule is either above the growth function, as the dashed curve through B, or it is below the growth function, as the dotted curve through A. There is a similar choice for initial conditions between d and G. In fact, the harvest rule is below the growth function to the left of d and above the growth function to the right of d.

We confirm this claim for points to the left of d using a proof by contradiction. The proof for points to the right of d mirrors the argument provided here. A proof by contradiction begins by hypothesizing the negation of the claim we want to establish, and then shows that this negation implies a contradiction; therefore the hypothesis is not correct. For example, if "Claim X" is either true or false, one way to establish that it is true begins with the hypothesis "Claim X is false." If we can show that this hypothesis implies something demonstrably false, then we conclude that the hypothesis is false; therefore, Claim X must be true. Here we want to show that for points to the left of d, the harvest rule lies below the growth function. Our hypothesis (the object we wish to falsify) states that for points to the left of d, the harvest rule lies above the growth function. We falsify this hypothesis by showing that it implies mutually contradictory results.

Under our hypothesis, the harvest rule intersects the origin, because harvest must be 0 if there are 0 fish. In addition, the harvest rule is continuous in x; if the harvest rule were discontinuous, there would be a jump in harvest at a point of discontinuity, and an associated jump in price. Such a jump violates the Euler equation, our no-intertemporal-arbitrage condition. For example, if there were a downward jump in harvest (e.g. harvest is bounded away from 0 at $x > 0$ and equal to 0 at $x = 0$), then there would be an upward jump in price; that could not be an equilibrium, because the owner would want to harvest less before the jump in order to increase harvest after the jump: there would be opportunity for intertemporal arbitrage. Clearly, for stocks near the origin, where the harvest is low, rent is positive.

Any curve (to the left of d) laying above the growth function (as our hypothesis states) and intersecting the origin (as we established in the previous paragraph) must be increasing in x over an interval near the origin; call such an interval J (merely to give it a name).[4] For $x \in J$, the stock is decreasing over time, because the harvest rule lies above the growth function. In addition, for $x \in J$, the harvest is decreasing over time, because the harvest changes in the same direction as the stock, which is decreasing over time. Because rent = price minus C, for $x \in J$ rent is increasing over time. In addition, for $x \in J$, $r < F'(x)$, because $F'(x)$ falls with x.

The previous two paragraphs establish that for $x \in J$: (i) rent is positive (ii) rent is rising over time, i.e. $\frac{dR}{dt} > 0$ and (iii) $r < F'(x)$. However, parts (ii) and (iii) and equation 16.1 imply that rent is falling over time, contradicting part (i). We have seen that our hypothesis ("The harvest rule lies above the growth function to the left of d") implies a contradiction. The hypothesis is therefore false. We conclude that the harvest rule must lie below the growth function to the left of d.

[4] In the interest of simplicity, Figure 16.2 shows the dashed curve through point B as increasing for all x. In this figure, the interval J equals $(0, G)$. Our argument does not require that the curve is monotonic over that entire interval, merely that there is a smaller interval over which it is monotonic.

16.3 Harvest rules for stock dependent costs

Objectives and skills

- Interpret graphs of the open access and the sole owner harvest rules.
- Use this figure to calculate the optimal open access tax.
- Introduce the phase portrait.

We carry out all of the analysis using the following parametric example[5]

$$\begin{aligned} c(x,y) &= \tfrac{C}{x}y \text{ with } C = 5 \\ F(x) &= \gamma x\left(1-\tfrac{x}{K}\right) \text{ with } K = 50 \text{ and } \gamma = 0.04 \\ p &= a - by,\ a = 3.5 \text{ and } b = 10, \text{ and } r = 0.03. \end{aligned} \tag{16.3}$$

We begin by summarizing the policy implications based on this example. Subsequent material develops the methods used to obtain those results. There, we explain the meaning and the use of the "phase portrait", the important new tool in this chapter. We then explain what a "full solution" means in a model of this sort.

16.3.1 Tax policy

Here we explain how to interpret Figure 16.3, and draw out its policy implications. For the model in equation 16.3, there is a unique steady state under the sole owner, $x_\infty = 39.35$, $y_\infty = 0.335$. We discuss only the behavior of the fishery for x below the steady state. Figure 16.3 shows the growth function, the heavy solid curve, for $x \le 39.35$. The dotted curve shows the harvest rule under the sole owner for x below the steady state. Most of the work involved with this analysis lies in identifying this harvest rule, i.e. in constructing the dotted graph. For the time being, we put those difficulties aside and discuss the meaning of this graph. (We do not use the thin

[5] An example with numerical values illustrates: (i) the insight that can be obtained from phase portrait analysis, and (ii) the recipe one follows in carrying out this analysis. The first item may encourage students to consult the texts listed at the end of this chapter in order to learn how to conduct this analysis; the second item may make that process easier, because students undertaking further study will have seen a motivating example.

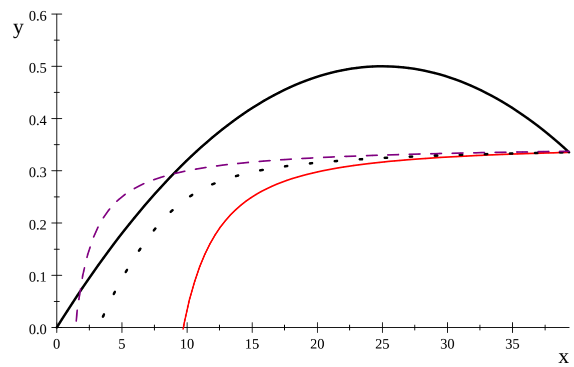

Figure 16.3: A part of the phase portrait, for stocks below the steady state level. The higher solid curve is the graph of the growth function. For combinations of stock and harvest, rent is zero on the dashed curve (the open access harvest rule). The dotted curve shows the sole owner harvest rule. The lower solid curve is used in the next section.

solid curve in the discussion here, although it plays an important role in the derivation below.)

The dashed curve shows the harvest rule under open access, obtained (as always) by setting rent = 0 and solving for harvest as a function of the stock. There are three steady states under open access, the points of intersection between the dashed and the heavy solid curve. The middle point ($x = 8$) is unstable, and the higher ($x = 39.3$) and the lower ($x = 2$) points of intersection are stable.

The harvest rule under the sole owner (the dotted curve) lies everywhere below the harvest rule under open access (the dashed curve). The sole owner always harvests less than the open access fishery. For stock levels close to $x = 39.35$, the sole owner harvests only *slightly* less than under open access; here it is not important to regulate the open access fishery. For stocks above but close to the open access unstable steady state ($x = 8$), harvest under open access is low enough to allow the fish stock to reach almost the optimal steady state; here, regulation allows the fish stock to recover more quickly, but it has no significant long term effect, because the steady states under open access and under the sole owner are almost the same. For stocks $2 < x < 8$, the stock declines to $x = 2$ under open access, whereas under the sole owner, it eventually recovers to 39.35. Over this range, regulation is important.

Given the two harvest rules, it is simple to determine the optimal tax, for any level of the stock. An example illustrates the procedure. Suppose that $x = 7$ (an arbitrary choice, merely for illustration). Reading from the harvest rules, we see that the open access harvest is approximately $y = 0.29$ and the sole owner harvest is approximately $y = 0.2$, yielding the equilibrium open access price $3.5 - 10\,(0.29) = 0.6$ and the equilibrium sole owner price $3.5 - 10\,(0.2) = 1.5$. The tax $\tau_{|x=7} = 1.5 - 0.6 = 0.9$ induces the open access to reduce harvest $y = 0.2$; at that level, the market price minus the tax equals the average harvest cost, and industry rent is zero. The tax $\tau = 0.9$ thus supports the efficient level of harvest at $x = 7$. Because the vertical distance between the two harvest functions changes with the level of the stock, the optimal tax also changes with the stock. The optimal tax is negligible for large stocks, but it is large at small stocks. For our example, the optimal tax comprises 60% of the equilibrium consumer price at $x = 7$. Recalling Chapter 15.2.1, the optimal tax under open access equals the rent under the sole owner. For this example, rent under the sole owner comprises about 60% of the market price when $x = 7$.

The tax implications of the models with constant and stock dependent costs are quite similar. In both, it is unimportant to tax the open access fishery at high stock levels. At high stock levels the optimal tax is zero under constant harvest costs, and the optimal tax is close to zero in our example of stock-dependent harvest costs. At low stocks, the tax is important in both models; it avoids physical extinction in one case, and economic irrelevance in the other. At intermediate stock levels, an open access tax can enable to fishery to recover more quickly, but has little or no long run effect.

Box 16.1 Back to Huxley and Gould Box 1.2 contains quotes from two 19th century figures, one explaining why regulation is not needed, and the other explaining why it is needed. The above analysis illustrates the circumstances where one or the other is correct. If stocks are above the unstable open access steady state, the open access outcome is at least approximately socially optimal. Stock dependent costs reinforce this tendency, by inducing fishers to reduce harvest as the stock falls. These forces provide a kind of automatic protection, as Huxley suggested, and there is little or no need for regulation. However, at low stocks, neither the market (which limits demand) nor technology (which limits supply by increasing costs) is adequate to protect the stock: regulation is needed, as Gould stated.

16.3.2 The phase portrait (*)

Equations 13.6 and 16.2 give the differential equations for the stock and for the rent under the sole owner. We can use these two equations, together with the definition of rent, $R = p(y) - \frac{C}{x}$, to obtain a third differential equation, for the harvest, $\frac{dy}{dt}$. It helps to give this differential equation a name, so we denote it as $\frac{dy}{dt} = H(x,y)$. (Appendix J.2 explains how we find this function H.) The solution to these three differential equations (in x, R, y) gives the optimal paths of the stock, rent, and harvest. Apart from the simplest problems, we cannot solve these equations analytically. Our first goal is to learn as much as possible about the solution without actually solving the equations: we seek qualitative information about the solution. The phase portrait is the key to achieving this.

The phase portrait contains two "isoclines". An isocline is a curve along which the time derivative of a variable is constant; here we set the constant to zero. Consider the logistic growth function, $\frac{dx}{dt} = \gamma x\left(1 - \frac{x}{K}\right) - y$. Setting this derivative equal to 0 gives $y = \gamma x\left(1 - \frac{x}{K}\right)$; the graph of this function is the x isocline (the curve where $\frac{dx}{dt} = 0$). Thus, the x isocline is simply the growth function; we are given that function as part of the statement of the problem, so no work is required to obtain the x isocline. We can also obtain the y isocline, the curve where $\frac{dy}{dt} = H(x,y) = 0$. Figure 16.4 shows the graphs of the two isoclines for our example. The intersection of these isoclines identifies the steady state, the point where $\frac{dx}{dt} = 0 = \frac{dy}{dt}$. For our example, the steady state is $x = 39.35$, $y = 0.335$.

In order to understand a phase portrait, the reader has to keep in mind that, outside the steady state, the stock and the harvest are changing over time. Imagine that there is a third axis, labelled time, t, perpendicular to the page, coming directly toward the reader. A point on the page represents a particular value of x and y at $t = 0$. A point above the page represents a particular value of x and y and a time $t > 0$.

Suppose that we start at $t = 0$, with some initial condition, $x_0 = x(0)$, and we pick some initial harvest, y_0. Starting from this point, there is a path, a curve in three dimensional space, call it (x_t, y_t, t) along which the differential equations $\frac{dx}{dt} = F(x) - y$ and $\frac{dy}{dt} = H(x,y)$ are satisfied. Now imagine shining a light from your eyes to the page. The curve (x_t, y_t, t), casts a shadow onto the page. We refer to this shadow as a trajectory. The phase portrait provides information about such a trajectory (i.e., the "shadow"); we use that information to infer facts about the behavior of the

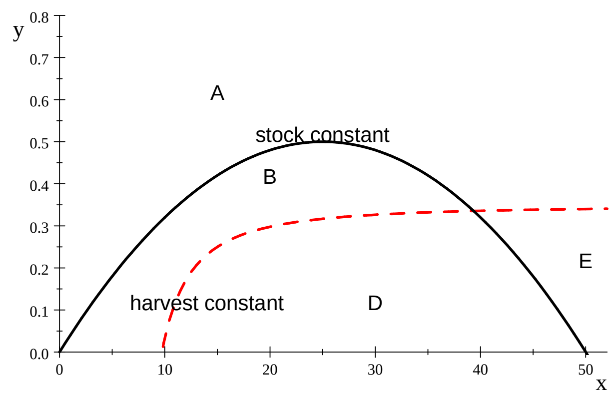

Figure 16.4: The solid curve: the x isocline (where $\frac{dx}{dt} = 0$; dashed curve: the y isocline (where $\frac{dy}{dt} = 0$). The two isoclines divide the plane into four regions, A, B, C and D, known as isosectors.

stock and the harvest over time. To this end, we use the four "isosectors" whose boundaries consist of the two isoclines. Figure 16.4 identifies these four regions as A, B, D, and E.

Isosector	A	B	D	E
motion of x	decreasing (west)	increasing (east)	increasing (east)	decreasing (west)
motion of y	increasing (north)	increasing (north)	decreasing (south)	decreasing (south)
overall motion of trajectory	north-west	north-east	south-east	south-west

Table 3. Direction of change of x and y and overall direction of motion of trajectory in the four iso-sectors

Consider isosectors B and D, the region below the x isocline. For any point in either of those two isosectors, $y < F(x)$. Consequently at such a point, $\frac{dx}{dt} = F(x) - y > 0$, i.e. x is increasing over time. For shorthand, we say that the trajectory is moving east (x is getting larger). Similarly, above the x isocline, in isosectors A and E, $y > F(x)$, so $\frac{dx}{dt} = F(x) - y < 0$. In these two isosectors, x is getting smaller, so we say that the trajectory is moving west. The second row of Table 3 summarizes this information.

We can identify the direction of movement in the north-south direction by using information about the differential equation for y, $\frac{dy}{dt} = H(x, y)$ (details in Appendix J.2). Below the y isocline (in isosectors D and E) $\frac{dy}{dt} = H(x, y) < 0$, i.e. y is decreasing, so the trajectory is moving south. Above the y isocline (in isosectors A and B) $\frac{dy}{dt} = H(x, y) > 0$, i.e. y is increasing, so the trajectory is moving north. The third row of Table 3 summarizes this information. The fourth row puts together the previous information, to obtain the overall direction of motion of a trajectory in each isosector.

This qualitative information tells us that an optimal trajectory that approaches the steady state x_∞ from a smaller value of x (to the west of x_∞), must lie in isosector B. Similarly, a trajectory that approaches the steady state from a larger value of x (to the east of x_∞) lies in isosector E. To explain and confirm these statements, we consider the case where the optimal trajectory approaches the steady state from below (i.e. from the west); the situation where the optimal trajectory approaches from above is similar.

A trajectory that approaches the steady state from below cannot lie in isosector E, because that isosector contains no stock levels less than the steady state. Therefore, the trajectory must lie in isosectors A, B, or D. The path cannot lie in isosector A, because trajectories there involve westward movements, i.e. reductions in the stock. Therefore, the trajectory must lie in either isosector B or D. It cannot lie in D, because from any point in D, it would be necessary for y to increase in order to reach the steady state; but trajectories in D move south, i.e. y falls there. Consequently, trajectories that approach the steady state from below (with stocks lower than the steady state level) do so in isosector E.

For this problem, regardless of the initial (positive) stock level, the optimally controlled fishery approaches the steady state. We can use this fact to establish that for any initial stock below the steady state, the optimal trajectory lies entirely in isosector B; and for any initial stock above the steady state, the optimal trajectory lies entirely in isosector E. For example, suppose that we begin with a stock below the steady state level. If, contrary to our claim, the trajectory beginning with this stock lay in either isosector A or D it would move away from the steady state.

Figure 16.5 repeats Figure 16.4, adding the dotted curve, the set of points where rent is zero. Rent is positive below the dotted curve and negative above it. We know that if the stock begins below the steady state, it ap-

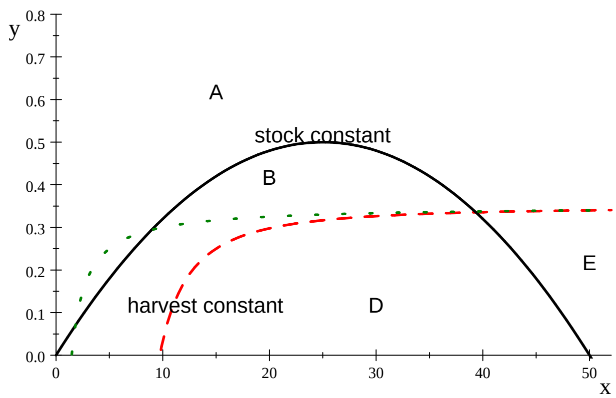

Figure 16.5: The dotted curve shows the combination of harvest and stock at which rent is 0.

proaches the steady state in isosector B, so the trajectory must be above the dashed curve (a boundary to isosector B). We also know that the sole owner never harvests where rent is negative, so the trajectory must be on or below the dotted curve. Therefore, we conclude that if the initial stock is below the steady state, the trajectory must be sandwiched between the dashed and the dotted curves. We obtained this information *without actually solving the optimization problem*, using only the necessary conditions for optimality and a bit of graphical analysis. This procedure illustrates the power of the phase portrait.

Given the resolution of Figure 16.5, it appears that the dotted and the dashed curves are coincident near the steady state. However, if we enlarged the figure in the neighborhood of the steady state, we would see that the dashed curve lies below the dotted curve. For this example, rent is always positive, but it is close to zero for fish stocks near the steady state. For different functional forms or parameter values, rent might be substantial along the optimal trajectory.

The full solution

Figure 16.3 contains the graph of the sole-owner's optimal harvest rule (the dotted curve in that figure). To construct that graph, we need the solution to the pair of differential equations $\frac{dx}{dt} = F(x) - y$ and $\frac{dy}{dt} = H(x, y)$ that includes the point (x_∞, y_∞), the steady state. The steady state is a "boundary condition" for this mathematical problem. As noted above, except for very few functional forms (not including our parametric example), there is no analytic solution to this problem. However, there are numerical routines that are straightforward to implement. The harvest rule shown in Figure 16.3 was obtained using Mupad, a feature of ScientificWorkplace.

16.4 Summary

Two examples, one with constant harvest costs, and the other with stock-dependent harvest costs, show how to analyze the dynamics under the sole owner. We explained how to determine the sole owner steady state(s), and to determine which steady state the sole owner fishery approaches, as a function of the initial condition. With constant harvest costs, this determination requires careful economic reasoning, but no new mathematical tools. With stock-dependent harvest costs, we require new tools. The most important of these is the phase portrait, which has many uses in dynamic problems.

Given this information, and using the fact that the open access fishery harvests up to the point where rent is zero, we were able to make qualitative statements about the optimal tax for the open access fishery. In particular, we learned that regulating this fishery is important if the stock is low; for sufficiently high stocks, regulation of the open access fishery is unimportant. This difference illustrates the general principle that for resource problems, the optimal tax is stock-dependent. In practice, we seldom have enough information to determine the optimal tax. A second best alternative is to close down the open access fishery at low stock, and leave the fishery untaxed at high stocks. A more nuanced policy imposes low or zero taxes at high stocks, and high taxes at low stocks.

Examples of this sort are useful for developing intuition. The pedagogic danger of these examples is that they may make the problem of regulation appear too simple. It might appear that all we need is a few parameter estimates and a modest knowledge of mathematics to propose optimal policy measures. That conclusion is too optimistic. The models studied here are good for the big picture, but they are too simple to be directly useful in actual policy environments. There, it may be important to consider multiple species or multiple cohorts of a single species, and different kinds of uncertainty, including uncertainty about functional forms and parameter values. Nevertheless, simple models provide a good place to begin.

16.5 Terms, study questions, and exercises

Terms and concepts

Continuous time Euler equation, isocline, phase portrait

Study questions

1. Using Figure 16.1, for each of the three values of C, sketch the sole owner's harvest rule that is consistent with the claims in Table 2.

2. (a) Using your sketch from #1 and $C = 3$, pick two values of x to the left of point f. At these two points, identify on the graph the differences in harvest under open access and under the sole owner. (b) Explain how you would use this graph to obtain the optimal tax under the sole owner, at these two values of x. (c) What qualitative statement can you make about the magnitude of the optimal taxes for the two values of x?

Exercises

1. By adapting the arguments used in Section 16.2.2, confirm the claims in the first row of Table 2 for x_0 to the right of point d.

2. By adapting the arguments used in Section 16.2.2, confirm the claims in the last two rows of Table 2.

3. Suppose that $F(x) = 0.04x\left(1 - \frac{x}{50}\right)$, inverse demand is $p(y) = 5 - 10y$, the discount rate is $r = 0.02$, and harvest costs are constant at $C = 0.4$. Suppose also that the initial condition, x_0, is below the horizontal coordinate of point e in Figure 16.1. (a) What tax (a number) supports an open access steady state at point d? (b) If the policymaker uses this constant tax, does it drive the stock to point d for all initial conditions below e? (c) For initial conditions below e (i.e., for values of x_0 below the horizontal coordinate of point e) does the optimal tax rise or fall with higher x?

4. For the example in the Exercise 3, suppose that $C = 3$. (a) Confirm that for positive initial conditions to the left of the unstable steady state (f), the sole owner drives the stock to point f. (b) Sketch the

harvest rule under the sole owner. (Your drawing will not be accurate, but you should be able to identify the points on the graph of the harvest rule, $y(x)$, corresponding to $x = f$ and $x = 0$. (c) Using this sketch, how does the first-best (stock-dependent) tax under open access vary with the stock, x?

5. For the model in equation 16.3, use Figure 16.3 to estimate the optimal tax for the open access fishery, at $x = 5$. Explain your steps.

Sources

Kamien and Schwartz (1991) provide many economic applications demonstrating the use of phase portrait analysis.

Clark (1996) uses phase portrait analysis for the fishery model.

Readers interested in extending the deterministic methods to a stochastic setting should consult Mangel (1985).

Chapter 17

Water Economics

Objectives

- Use the tools developed in previous chapters to study other natural resource problems.

Skills

- Be familiar with market failures associated with water.
- Use both static and dynamic methods to study water problems, and to analyze policy remedies.

We used oil and fish in developing analytic tools to study nonrenewable and renewable resources, and also to illustrate market failures and appropriate policies. This chapter introduces water economics, emphasizing the generality of both the policy problems and the tools discussed in previous chapters. Nonrenewable resources, like oil, do not regenerate on a time-scale relevant for human planning. Renewable resources, like fish, potentially regenerate quickly, over a period of years or decades. Water, forests, and many other resources, are intermediate cases. Water in a slowly recharging aquifer (a geological formation that stores water) and the stock of old growth redwood trees are, for practical purposes, nonrenewable resources. Water in a lake (with inflows) and new-growth forests are renewable resources.

Water is an essential, and in many parts of the world, poorly managed natural resource. A 2012 US government "Intelligence Community Assessment" anticipates that during the next decade many countries will experience problems caused by water shortages, poor water quality, and floods.

These problems contribute to political instability and regional tensions. Absent policy changes, growing water demand will outstrip supply, jeopardizing production of food and energy, and putting at risk economic growth. Only about 2.5% of the earth's water is freshwater. Glaciers contain about 69% of the freshwater and groundwater (water in aquifers) about 30%. The surface (rivers, lakes) and atmosphere contain about 0.4% of freshwater.

Most uses of water have both "consumptive" and "nonconsumptive" uses. Hydroelectric power generation does not reduce the quantity of water, and is therefore considered a nonconsumptive use. However, the dams constructed to create hydroelectric power reduce the availability of water for fish runs and other environmental or recreational purposes. Dams can also reduce water quality due to the buildup of silt and increased salinity. Much of the water used for agricultural irrigation is absorbed by plants and the atmosphere, a consumptive use, but some of it returns to rivers and aquifers, a nonconsumptive use. However, when agricultural runoff is polluted, these return flows create costs, not benefits. Agriculture accounts for about 78% of total (consumptive plus nonconsumptive) use, with household and industrial (19%) and power generation (10%) making up most of the remainder Agriculture accounts for about 93% of consumptive water use.

A larger and more prosperous population increases the demand for water, as more people eat a more water-intensive diet. Improved technology and additional infrastructure can help to offset the growing imbalance between water supply and demand. Drought resistant crops and the development and adoption of more efficient irrigation can reduce the amount of water needed to grow a given amount of food; infrastructure investments can reduce waste by reducing leaks. Technical remedies are important in solving a global water shortage, but without policy changes they are unlikely to be adequate.

In many places, political decisions impede obvious solutions and create perverse incentives that make water problems worse. We provide examples of these, and then discuss static and dynamic market failures. A static market failure results in the inefficient use of a given flow of water. A dynamic market failure results in too rapid use of water resources.

17.1 The policy context

Current water laws and policies result from the accretion of decades, and in some places centuries, of social interactions. Not surprisingly, in many cases these policies are inefficient. This section discusses two types of policy failure: (i) water is priced inefficiently, or not at all; (ii) policies not directly targeted to water use make water problems worse.

The efficient use of water creates positive "water rents", just as the efficient extraction of oil or harvesting of a fishery lead to positive resource rents. These rents equal the opportunity cost of water use, arising from water's scarcity and from higher future pumping costs. Efficient management of water resources requires that the user price includes not only the cost of providing (pumping and transporting) water, but also rent, the opportunity cost of the water. A price that includes only the cost of providing the water (i.e., excludes rent) is too low, and leads to excessive water use. We consume water as a "bundle", consisting of the liquid itself, and its location at a point in time. Putting aside contamination that might arise during transportation, the physical object is the same if it exists in our kitchen tap or in an aquifer hundreds of miles away. If we pay only for the cost of transporting the water from the aquifer to our kitchen, without paying rent, then we are paying only part of the real cost of consuming the water.

If users pay a single price per unit that includes rent, sellers' revenue exceeds their cost of provision. A California law forbids municipalities from charging more for utilities (e.g. water) than the cost of provision. If the "cost of provision" is narrowly construed, to exclude resource rent, this law makes efficient pricing impossible. "Tiered pricing", which allows the price to vary with the amount consumed, enables a municipality to induce the optimal level of water consumption while not earning profits (or rent), and simultaneously providing a subsidy to people who use little water (typically, poorer people).[1] Efficient tiered pricing requires that the marginal (highest) use be charged at the efficient price, equal to the marginal cost of provision plus the rent. Lower quantities can be charged at lower prices, even below the marginal cost of provision. Under this structure, the utility makes profits from selling to high-quantity users, using those profits to subsidize low-quantity users.

Figure 17.1 shows an aggregate demand function and a supply function with constant pumping + transport marginal = average cost, $C = 3$.[2] Sup-

[1] A 2015 California State Supreme Court ruling upheld an appellate court's decision that struck down a municaplity's use of tiered pricing. The appellate court did not ban tiered pricing, but required that its structure be tied to the cost of providing service.

[2] The aggregate demand function is the horizontal summation of the individual household demand functions, which are not shown. Those demand functions vary with income and other household characteristics.

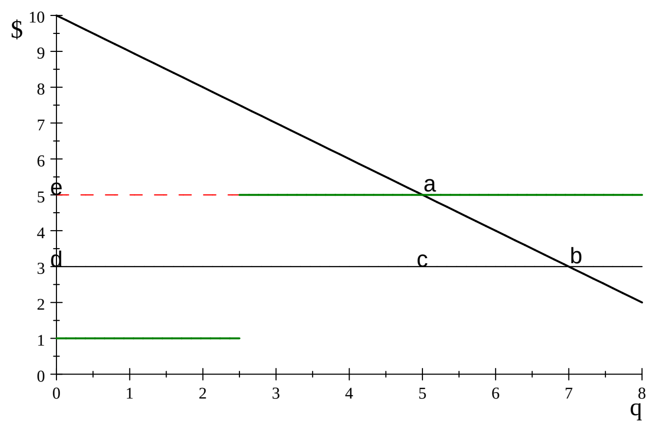

Figure 17.1: Cost of providing water = 3 and optimal rent = 2. The loss in consumer welfare from moving to efficient water pricing under a single price is the trapezoid *abde*. The loss in consumer surplus under tiered pricing, where consumers obtain all of the resource rent, is the triangle *abc*

pose that the socially optimal level of rent is $R = 2$. A municipality that charges a single price greater than 3 has revenue exceeding costs, earning profits. A municipality that charges a single price less than 5 induces socially excessive water use. By using a price less than 3 for consumption below a threshold, and using a price equal to 5 for consumption above that threshold, the municipality can achieve the optimal level of water consumption and break even. Figure 17.1 shows the case where the low price is 1 and the low-price threshold is 2.5.

Tiered pricing can reduce the water bills of low-use consumers. High-use consumers who see their water bills rise are likely to be wealthier. For the example in Figure 17.1, moving from the inefficient price $p = 3$ to the efficient single price, $p = 5$, reduces consumer surplus by *abde*. In contrast, moving to efficient tiered pricing with zero profits for the municipality, reduces consumer surplus by *abc*. Whenever the socially optimal rent is positive, efficiency reduces aggregate consumption (from 7 to 5 in this example), lowering aggregate consumer surplus. Tiered pricing can reduce the fall in aggregate consumer surplus from the large trapezoid to the small triangle, while transferring welfare from the rich to the poor.

By 2015, over half of California's water districts used some form of tiered pricing; at the same time, in many Californian communities, water was not even metered. In Riverside California, tiered pricing reduced water demand by 10- 15%. Santa Fe, New Mexico used tiered pricing with high marginal prices, and had a per capita consumption of about 100 gallons per day; Fresno, California, with a low uniform water price, had a per capita consumption of over 220 gallons per day.

Laws that impede or prohibit efficient pricing are perhaps the most obvious examples of policy failure. However, many policies ostensibly unrelated to water have major implications for water use. Examples illustrate this relation. Strong U.S. sugar lobbies have propped up domestic sugar prices by maintaining restrictions on U.S. imports of lower-cost foreign sugar. This method of supporting U.S. producers is politically attractive, because, unlike direct subsidies (which have been widely used for export crops such as corn) the trade restrictions have no direct budgetary costs; consumers, not taxpayers pay for the subsidy to producers. The high domestic prices encourage domestic production, which in the case of sugar has led to wasteful use of water and associated pollution in the Florida Everglades. Chapter 9.5 shows that output and input subsidies tend to reinforce each other: a positive output subsidy can greatly increase the welfare loss arising from under-priced natural resource inputs. Elsewhere, U.S. subsidies have promoted the production of water-intensive crops in drought-prone areas, e.g. rice production in California. U.S. ethanol policy, implemented by the Renewable Fuel Standard (Chapter 9.3), is an indirect subsidy to corn producers. This subsidy has encouraged irrigated corn production in the high plains, adding pressure to the Ogallala aquifer (Chapter 17.3).

Similar problems arise in many parts of the developing world. The Zayanderud River, which runs through the Iranian city of Isfahan (population of 2.5 million) went dry in the early 2010's, while groundwater levels fell and wells dried up. A drought that began in 1999 and worsened in 2008 precipitated the crisis, and mismanagement exacerbated it. As part of political maneuvering to increase local support, the central government transferred control of the watershed from a unified authority to local leaders, who then allocated water without regard to the resource constraint. Crop subsidies that increased the demand for water, and local leaders' support for water intensive industries, worsened the problem. Popular sentiment opposed rational water pricing.

In India, tube wells increased irrigation in the Ganges watershed. In the 1980s, Uttar Pradesh subsidized the cost of well construction, and banks extended credit for pumps. Users' low electricity price encouraged pumping, and landowners were not charged for groundwater extraction. Over-extraction caused water tables to fall, increasing pumping costs and endangering public hand pumps used primarily by the landless poor. India's Ground Water Authority banned private extraction and sale of groundwater in some areas, but illegal pumping continued. A 2012 scientific report stated that the aquifer that serves the capital, New Delhi, could dry up in a few years.

Examples of these sorts can be found in many countries. Increases in water demand, due to higher population and higher living standards, put pressure on limited resources. In some cases, these supplies are further stressed by droughts and pollution. New infrastructure, in the form of dams, aqueducts, and replacement of leaky pipes, and better technology, in the form of more efficient irrigation and desalinization plants, can help to solve or at least postpone crises. Rationalizing water pricing and reforming policies that worsen water shortages, can make the problem more tractable.

17.2 The static market failure

In many parts of the world, including eastern U.S. states, water rights are based on riparian (pertaining to riverbanks or wetlands) law; landowners have the right to use water on their land, provided that their use does not conflict with other riparian users. Typically, this use does not include irrigation. In the western U.S. states, water rights arise from "prior appropriation", having been the first to make "beneficial use" (e.g. irrigation) of unclaimed water. This basis for water rights led to a rush of sometimes fraudulent water claims in the west. It also created the incentive to use water, sometimes inefficiently, partly to forestall others from making their claim. In the first half of the 20th century, western U.S. states rushed to build dams and irrigation projects. These property rights entitle owners (individual farmers or cities or states) to use, but not to sell "their" water.

Here we take the aggregate supply of water as fixed within a period, and discuss its allocation across uses. The "static market failure" arises from the inefficient use of a given flow of water. Water owners' inability to trade their water creates this static market failure. We illustrate this graphically, explain some of its causes, and then consider a political economy implication. Figure 17.2 shows the inverse demand functions for two agents (solid and dashed) when the aggregate supply is $y = 5$. If one agent is a

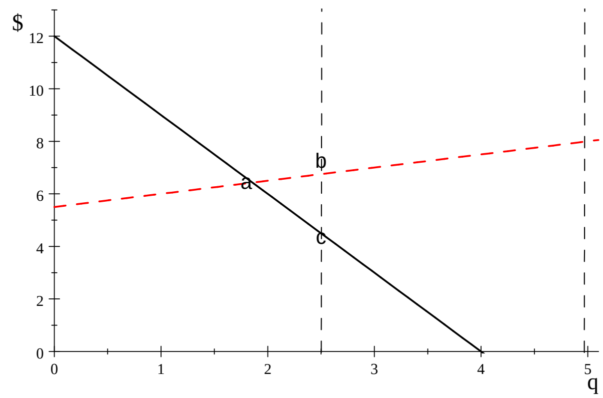

Figure 17.2: Two water demand functions, $p = 12 - 3q$ (solid) and $p = 8 - 0.5\,(y - q)$ (dashed) when the aggregate supply is $y = 5$. Consumption of the "solid" firm is q and of the “dashed” firm is $y - q$. The efficient allocation occurs at $q = 1.9$, where both firms have the price $p = 6.4$. If firms have equal shares of total supply, $q = \frac{y}{2}$, then they have different marginal willingness to pay, at points b and c.

household, then the inverse demand function has the usual interpretation, as the marginal willingness to pay for an additional unit. If an agent is a firm (or farm), the inverse demand is the value of marginal productivity of water: the additional value to the firm created by using an additional unit of water. Figure 17.2 looks like Figure 2.1. From the discussion of arbitrage in Chapter 2.1, we know that an efficient outcome occurs where each agent has the same marginal willingness to pay, at point a; there, the “solid” agent consumes $q = 1.9$ units and the “dashed” agent consumes $5-1.9 = 3.1$ units.

To explain the static inefficiency, suppose that both users have property rights to half of the total allocation ($y = 5$), but an institutional constraint prevents them from trading. They each consume 2.5 units, and have the willingness to pay shown by points b and c. The prohibition against trade causes a welfare loss equal to the area of the triangle abc. The “dashed” agent would be willing to pay $p = 6.75$ for an extra unit of water, and the “solid” agent would be willing to sell a unit of water for $p = 4.5$. At the constrained outcome there are potential, but unrealized, gains from trade.

Figure 17.3 graphs, as a function of y, the welfare loss (the area of the triangle abc). The solid curve shows this loss as a percent of the welfare level in the constrained scenario, in which each agent consumes half of the

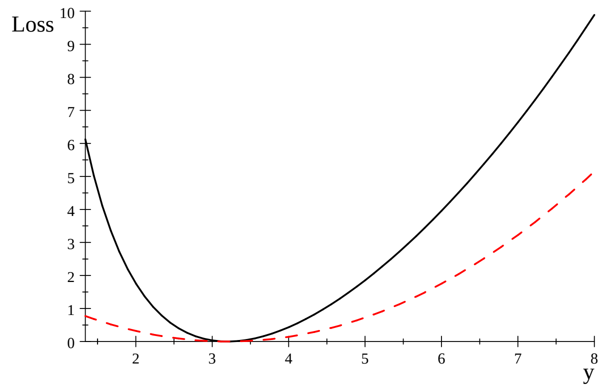

Figure 17.3: The solid curve shows the welfare loss, as a percent of welfare under the constaint $q = \frac{1}{2}y$, as a function of aggregate supply, y. The dashed curve shows the absolute welfare loss.

available supply; the dashed curve shows the absolute welfare loss. There is a particular value of y (equal to 3.2 for this example) at which it is optimal for each firm to consume the same amount. At that point, the welfare cost of the constraint is 0. For levels of y close to 3.2, the welfare loss due to the constraint is small. However, it begins to rise quickly as y moves away from 3.2: the welfare cost is convex in the aggregate supply of water.[3]

The static market failure arises from a missing market: the inability of the agents to trade their endowment. Water economists have devoted substantial effort over the last several decades measuring and explaining the welfare consequences of this policy-imposed market failure. Agents' inability to trade their endowment is a type of imperfect property rights: agents can use but not sell their property.

Reasons for the prohibition on trade

In many cases, the historical evolution of property rights, not conscious design, explains the prohibition against water trades. This limitation is sometimes justified on the grounds of fairness, based on the idea that water is a gift on nature: people should be allowed to use, but not to own this part

[3]Given the parameter values in this example, we require $1.34 \leq y \leq 8$. For $y < 1.34$ the "dashed" agent has negative consumption under efficient allocation, and for $y > 8$ the "solid" agent has negative price under the inefficient allocation.

of nature. This view is inconsistent with the fact that people own land and mineral rights, which are equally gifts of nature. A related fairness argument recognizes that in many areas the value of water stems from previous social policy, not just from nature's largess. In western U.S. states, as in many parts of the world, publicly funded water projects created dams to store water and aqueducts to transfer it to (principally) farmers. Most of these projects were originally intended to be funded by the water users, but in practice they were heavily subsidized by taxpayers. (An example in Chapter 2.5 shows that an extended no-interest loan can amount to a sizeable implicit subsidy.)

Current farmers bought or inherited land, and with it, the attached water rights. The increase in the value of water, resulting from allowing transfers, provides a windfall to those with water rights. The people who would be able to buy the water also gain: both buyers and sellers gain from a transaction. By taxing some of the surplus created by water transfers, and using it for public objectives, the gain can be spread more broadly.

Water transfers benefit both buyers and sellers, but they may harm third parties. Some agricultural water use is consumptive, and some is nonconsumptive, because some irrigation water returns to rivers and aquifers, where it can benefit other users. By removing water from the hydrologic system, e.g. transferring it from agricultural to urban use, the trade can harm third parties who would otherwise benefit from the recycled water. Water transfers can also harm workers or businesses that benefit from a strong local economy. If the transfer occurs from agriculture to urban use (as has recently been the case in western U.S.), the demand for farm labor can fall, lowering the wage or employment. The reduction in agricultural output can also decrease demand for local farm services (e.g. machine sales and maintenance). These third parties may have a strong incentive to block transfers.

These externalities or other market failures might cause the transfer-induced reduction in third party benefits to exceed the direct welfare gain arising from the transfer. In that case, allowing water transfers, without correcting the other market failures, lowers welfare. Second-best arguments have to be examined critically, on a case-by-case basis, because these kinds of arguments can be constructed to oppose almost any reform. Even if the third party argument does not provide an efficiency rationale for prohibiting water transfers, it helps explain the resistance to those transfers. The creation of water markets requires consideration of third party consequences.

All of these issues are present in fisheries, and in many other resource settings. In the fishery context, we noted the importance of creating individual *transferable* quotas (ITQs, not just IQs). The transferability enables the market to reallocate quotas to the most efficient fishers. This reallocation creates a welfare gain, just as does the reallocation of water in the example shown in Figure 17.2. There are also third-party issues in the fishing context, related to local fishing communities. Although each natural resource gives rise to specific problems, the different resources share many of the same features. Thus, the skills and intuition acquired in studying one type of resource often help in studying a different resource.

Political economy implications

We use Figures 17.2 and 17.3 to make a general point about political economy. The fact that efficiency increases the size of the economic pie, might suggest that greater efficiency makes it easier to reach an agreement amongst competing interests. However, people's incentive to influence political decisions, i.e. to maintain or increase their water allocation, likely depends on the value of an *additional* unit of water, not only on the value of the water they currently own. The distinction is between the value of water and the marginal value of water. These two objects do not necessarily move in the same direction. The marginal value of water, equal to the amount that people would pay for an extra unit of water, is the inverse demand function for water. An institutional change, such as opening water markets, typically increases the value of water, but it might either increase or decrease the demand for water.

Suppose that three interest groups, agricultural users, urban users, and environmentalists, compete for water; initially, there are no water markets. Figure 17.2 shows the water demand from agricultural and urban users. The area under the demand function for urban users equals their consumer surplus; the area under the value of marginal product curve for the agricultural users equals their profit from water use. The aggregate area under the curves equals the combined value of their allocation. Denote this combined area, when transfers are not allowed, as $V(y)$; it depends on the total allocation, y, and on the split (one half for each user in this example). The marginal value, $\frac{dV(y)}{dy} = V'(y)$, is the increase in these users' combined value, due to an extra unit of water, given that they share this extra unit equally; $V'(y)$ is the aggregate demand function for urban and agricultural water users under the equal-sharing constraint.

Denote $V^{opt}(y)$ ("opt" for "optimal") as the combined payoff of urban

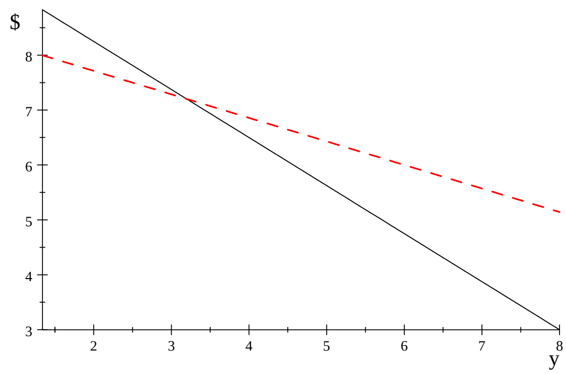

Figure 17.4: The solid line shows the aggregate demand for water ($\frac{dV}{dy}$) under the constraint $q = \frac{y}{2}$. The dashed line shows the aggregate demand for water ($\frac{d(V+\Delta)}{dy}$) when users can trade their allocations. Water markets increase the demand for water if and only if $y > 3.2$.

and agricultural users when a water market permits transfers between them. The loss in surplus arising from the constraint that prohibits water transfers (the area of the triangle *abc* in Figure 17.2) is $\Delta(y) \equiv V^{opt}(y) - V(y)$. This loss is positive, except for the knife-edge value of y (3.2 in the example). The constraint unambiguously lowers the *value* of water. But how does the constraint affect the *demand* for water? The model provides a simple answer. The inverse demand under the constraint is $D(y) = \frac{dV(y)}{dy}$, and the inverse demand under water transfers equals

$$D^{opt}(y) = \frac{dV^{opt}(y)}{dy} = \frac{d(V(y) + \Delta(y))}{dy} = D^{con} + \Delta'(y). \tag{17.1}$$

Figure 17.3 shows that $\Delta'(y)$ is positive or negative, depending on the magnitude of y.

Figure 17.4 shows the aggregate consumptive demand functions without (solid) and with (dashed) water markets. For this example, liberalizing markets (allowing water transfers) increases water demand if and only if $y > 3.2$.[4] In general, a reform that moves us closer to efficiency increases the *value* of

[4] This conclusion is based on a simple example, but the geometry shows that it relies only on the shape of the loss function $\Delta(y)$. Under quite general circumstances, this function is convex and reaches a minimum at an interior value. Under these circumstances, the result described in the text holds. Often examples can reveal general insights.

water. However, there is no presumption that this reform increases the *demand* for water. A change in institutions (or technology, or policy) can have qualitatively different effects on the function V and on its derivative V'.

The urban, agricultural and environmental lobbies engage in the political process that determines the allocation of a fixed flow of water. The amount available for environmental services (e.g. fish spawning) equals this flow minus the aggregate allocation for urban and water users, y. In order to make our point simply, we assume that urban and agricultural interests take as given their split, so the only way that either can increase their own allocation is by increasing the combined allocation, y. With this assumption, urban and agricultural users are in a natural alliance. An increased allocation to urban and agricultural users comes at the expense of environmental interests, who in this example are arrayed against the urban–agricultural alliance.

Suppose that we begin in a scenario without water markets. The urban-agriculture allocation, y, is determined by a political process that (along with physical and technological considerations) balances the urban-agriculture and environmental interests. The political economy forces depend in part on the lobbying effort of the agricultural and urban users, both of whom want to increase y (because $V'(y) > 0$). Their lobbying effort, and consequently the (political) equilibrium value of y, depends in part on the value to the consumptive users of an *additional* unit of water.

Now an economist enters the debate, promoting the adoption of water markets on efficiency grounds. Should an environmentalist support or oppose this institutional reform? The previous analysis shows that the adoption of water markets might either increase or decrease the consumptive users' marginal value of water, thereby increasing or decreasing their incentives to lobby for a higher allocation. If it increases their incentives to lobby (i.e. if initially $y > 3.2$ in our example), then the institutional reform likely increases the political opposition that the environmentalist faces. Even in this case, the environmentalist might support the reform as a part of a package deal that also protects the environment, e.g. by decreasing or at least not increasing the consumptive allocation, y. Because the introduction of water markets increases the consumptive users' value of a given allocation, rational users would agree to a reduction in y in exchange for a reform that increases the value of y.

A move to efficiency creates surplus that can be used to compensate environmentalists (or other third parties). Neither markets nor political economy interactions automatically lead to this compensation. In our example, the political economy market, operating through lobbying, can lead to a worse environmental outcome after an institutional reform that allows water transfers. Political bargains potentially enable environmentalists to support institutional reform while also protecting the environment.

One further point deserves mention. It is sometimes assumed that changes that increase the value of a resource tend to reduce resource demand. For example, better technology, in the form of drought resistant crops or more efficient irrigation techniques, make it possible to achieve the same level of production with less water. These changes might also increase water demand, by making it profitable to grow crops that were previously uneconomical. This increase illustrates the "rebound effect", the situation where a change that would appear to decrease demand for a resource, ends up increasing demand. Again, this possibility arises because demand depends on the marginal value of an additional unit of a resource. A change can unambiguously increase the value of the resource, while having an ambiguous effect on the marginal value of the resource.

17.3 The dynamic market failure

The static question is "How should we use a given amount of water (a flow) in a period?" The dynamic question is "How should we manage a given stock of water, i.e. choose the flow trajectory?" Full efficiency requires that water allocations be arbitraged over competing users within a period, and that they be arbitraged across time. Both the static and dynamic market failures arise from imperfect property rights, which interfere with intra- and inter-temporal arbitrage.

If all of our water came from (free flowing) rivers or from annual rainfall, then the policy problem would be static. Nature would determine the availability of water in each year, and policy would determine the allocation of that water across competing uses. Dynamics are important, because much of our water supply is stored in reservoirs, lakes, and aquifers. The Ogallala aquifer illustrates dynamic water problems. We then provide an analytic foundation, building on resource models from previous chapters.

17.3.1 The Ogallala aquifer

The Ogallala Aquifer, located beneath eight U.S. states from Texas to South Dakota, exemplifies the problem of managing a common property resource. This million-year old aquifer, ranging over 174,000 square miles, provides water for almost a fifth of U.S. wheat, corn, cotton and cattle production; agriculture accounts for about 95% of water use. The aquifer contains enough water to cover all 50 states with 1.5 feet of water; if it went dry, it would take natural processes 6,000 years to refill. Extraction during the first decade of the 21st century was a third of total extraction during the previous century. The stock of water in the aquifer declined 10% from the early 20th to the early 21st century. Water levels in 25% of the land above the aquifer fell by over 10 feet. There may be enough water in northern regions to last hundreds of years, while in the southern High Plains a third of farm land may lose irrigation over the next several decades.

Withdrawals from the Ogallala, made economical by the introduction of the center-pivot irrigator, accelerated in the 1940s and 50s. Technological advances, including more efficient irrigation or drought resistant crop varieties, *might* reduce water demand. (Keep in mind the rebound effect, described above). However, in most locations sustainable use of the aquifer requires lower withdrawals. Some reductions might be accomplished by switching to less water intensive crops (e.g. sunflowers instead of corn), by changing cultivation practices (e.g. adopting "no-till" methods), or retiring land from cultivation. These changes require short run sacrifices, which are hard to enforce when decisions are made by thousands of farmers in a common property setting.

The Ogallala is nominally a regulated resource, with rules varying across states. Nebraska passed laws in the 1970s limiting water allocations and using rotating water permits. Despite enforcement problems, Nebraska has been successful in maintaining groundwater supplies. Elsewhere, regulation has not prevented rapid declines in the aquifer. By the 1970s, the fact that the Ogallala is a finite resource was widely recognized; in the mid 1980s, heads of water conservation boards in Colorado and New Mexico stated that their goal was to make this resource last for 25 – 50 years. With this objective, it is not surprising that the aquifer is being depleted rapidly. In Texas, regulation largely consists of restrictions on the distance between wells and from wells to property lines. Of the nearly 100 Texan groundwater conservation districts, the Texas High Plains district was one of the first to limit the amount of water pumped. Their goal was to conserve half of the stock available in 2010 until 2060. A Texas Supreme Court 2012 opinion delayed the execution of this ruling, questioning its legality on the basis that landowners have the same property rights to the water beneath their land as to the oil and gas.

Kansas law allows conservation districts to limit water withdrawals, but no such limitations were imposed from 2009–2014, despite falling water levels. State officials claimed that mandating lower withdrawals would be heavy-handed. Kansas law enables farmers to create groups that, with a two-thirds vote, can restrict water withdrawals for all farmers in the area. This plan reduces the transactions cost that arise when thousands of farmers over vast expanses have to reach an agreement. Like the marketing orders described in Chapter 9.3, it provides a possible remedy to a common action problem: here, conserving water. Two years after Kansas made these associations legal, only one group of 110 farmers formed such an agreement. Geological factors add to the usual problems in getting a group with competing interests to cooperate on a mutually beneficial plan. There is considerable geographical variation in the aquifer's lateral permeability. A successful farmer group must include a large enough area to insure that little of the water saved by the group migrates to parts of the aquifer below land owned by non-members. Otherwise, the group's conservation largely benefits non-members. Farmers near the boundary of the group are likely to face increased movement of water outside the group boundary. These geological factors complicate the problem of managing water by means of voluntary groups.

17.3.2 A model of water economics

To examine the dynamic inefficiencies, we abstract from the complexities associated with a particular aquifer. We also assume away the static inefficiency discussed in Section 17.2. For example, all of the users might have the same marginal value of water and an equal allocation, so they have no motivation to trade; alternatively, they may be allowed to trade amongst themselves, so that in equilibrium they have the same marginal value. As above, we denote the aggregate value of a given flow y as $V(y)$.

The lessons learned from fishery economics, in particular the distinction between open access and private ownership, are relevant in the water context. In the fishery setting, absent regulation, open access (free entry) drives equilibrium rent to zero. In the agricultural setting, only landholders can pump water, creating a barrier to entry, and leading to the possibility of positive water rents even when water extraction is unregulated. However, those rents are likely to be small: an individual landowner has little (selfish) incentive to conserve the stock of water. Conservation reduces short run profits; because the aquifer is porous, much of the saved water migrates to neighbors' land, where others would use it in the future. Therefore, the outcome under common property, with many users, is similar to the outcome under open access.

We abstract from the fact that the aquifer is not perfectly porous, assuming that all users draw from a single stock of water. The model does not describe the entire Ogallala aquifer, but it can describe a region below which lateral movement of water occurs quickly, e.g. over a period of years, not centuries. We refer to this region as the aquifer. Retaining notation from previous chapters, we use x_t to denote the stock of water in the aquifer at the beginning of period t, and y_t to denote the amount of water taken from the aquifer in period t. The change in the stock is

$$x_{t+1} - x_t = F(x_t) - y_t. \tag{17.2}$$

If $F(x) \equiv 0$, we have the nonrenewable resource model of Chapter 5. For $F(x) \neq 0$, we have the fishing model of Chapter 13. In simplest case, where $F(x_t) = \alpha$, a constant, there is an exogenous flow of water into (if $\alpha > 0$) or out of (for $\alpha < 0$) the aquifer. More generally, the amount of water currently in the aquifer, x_t, might affect current and future flows. If the stock of water falls below a critical level, land above the aquifer may subside, reducing the aquifer's ability to store water. The flow also depends on x_t if our aquifer is part of a larger hydrologic system, and the stock of water in neighboring parts of the system is exogenous; those stocks might be under a different regulatory regime. The net flows to our aquifer depend on the relative pressure in the different parts of the hydrologic system; the relative pressure depends on x_t and on the exogenous stock outside our aquifer.

The cost of extracting and transporting y units of water when the stock is x equals $(c_0 - cx)\,y$. For $c > 0$, a larger stock reduces these costs. Pumping costs are lower when the water table is higher, corresponding to a larger stock of water. There are two aspects of the externality associated with common property aquifers: (i) Increased pumping reduces the stock of water, making less available for other users in the future. (ii) The lower stock of water also increases other farmers' future pumping costs. Scarcity and extraction costs were important in our discussion of rent for both the nonrenewable resource (Chapter 5) and the renewable resource (Chapter 15).

Agents' incentives to extract water depends on the relation between the marginal utility and marginal cost of water extraction, $V'(y)$ and $(c_0 - cx)$. An increase in $V'(y)$ increases the incentive to extract, and a decrease in x makes extraction more expensive, decreasing the incentive to extract. We study extraction decisions under a social planner (perfect regulation) and then under common property (no regulation). Extraction from aquifers lies somewhere between these two extremes, but in many regions is closer to common property.

The social planner

As in previous chapters, we denote the discount factor as ρ, and write the present discounted value of the stream of water use, $\{y_0, y_1, y_2, ...\}$, as

$$\sum_{t=0}^{T} \rho^t \left[V(y_t) - (c_0 - cx_t)\, y_t\right], \tag{17.3}$$

where T is the last period during which extraction is positive. Depending on the nature of regulation and the parameters of the model, extraction might continue indefinitely ($T = \infty$) or end in finite time. We can use the perturbation method (Appendix I) to write the Euler equation under a social planner (first-best regulation) as

$$V'(y_t) - (c_0 - cx_t) =$$
$$\rho\left[\underline{\left(V'(y_{t+t}) - (c_0 - cx_{t+t})\right)}\underbrace{\left(1 + \frac{dF(x_{t+1})}{dx_{t+1}}\right)} + \underline{\underline{cy_{t+1}}}\right]. \tag{17.4}$$

Equation 17.4 has the same interpretation as in the fishery setting. We can perturb a candidate trajectory by extracting one more unit of water in the current period, and making an offsetting change in the subsequent period, so that x_{t+2}, the stock in the next period, equals the level under the candidate trajectory. If the candidate is optimal, the marginal gain from this perturbation should exactly equal its marginal loss, so that the perturbation has no (first order) effect on welfare.

The term on the left side of equation 17.4 is the marginal gain due to extracting an additional unit of water in period t, the difference between the marginal benefit and the marginal cost. In order to return the stock to the candidate level at $t+2$, we must reduce extraction at $t+1$ by the under-bracketed term. (See Chapter 15.1.1.) Each reduction in next period extraction reduces benefits by the single-underlined term. The reduced benefit caused by the lower extraction at $t+1$ equals the product of these two terms. In addition, the lower stock at period $t+1$ increases extraction costs by the double-underlined term. Thus, the right side of equation 17.4 equals the present value of the marginal loss of the perturbation. The Euler equation states that if the candidate is optimal, the marginal gain from a perturbation must equal the marginal loss of the perturbation.

In previous chapters, where the single period payoff equals revenue minus costs, we defined rent as marginal revenue (= price for the competitive firm) minus marginal cost. Here, the single period payoff equals $V(y_t) - (c_0 - cx_t)\,y_t$, the current benefit minus the cost of extracting y; we accordingly define water rent as marginal benefit minus marginal cost:

$$R_t = V'(y_t) - (c_0 - cx_t)\,. \tag{17.5}$$

Using equation 17.5, we simplify equation 17.4 to obtain

$$R_t = \rho\left(R_{t+1}\left(1 + \frac{dF(x_{t+1})}{dx_{t+1}}\right) + cy_{t+1}\right). \tag{17.6}$$

Regulation (potentially) leads to a high steady state stock only if the planner cares enough about the future (has a high discount factor). A social planner with a low discount factor might drive the stock to a low level, or exhaust the aquifer. Figure 17.5 shows the steady state stock of water (solid) and a multiple of the extraction (dashed), as a function of the discount factor, ρ.[5] A larger discount factor implies that the planner cares more about the future, leading to a higher steady state stock of water. The extraction, y, in contrast, is non-monotonic in the discount factor. For $\rho < 0.93$, the higher stock corresponding to a higher discount factor makes it possible to extract more in the steady state. However, for discount factors above $\rho > 0.93$, it is necessary to decrease steady state extraction in order to maintain a higher steady state stock. The planner with high ρ is willing to decrease extraction

[5] This example uses benefit $= V = 10y - \frac{1}{2}y^2$, cost $= (c_0 - cx)\,y = (10 - 5x)\,y$, and growth $= F(x) = 1 + 0.1x\left(1 - \frac{x}{100}\right)$. This example has an exogenous (stock independent) recharge rate $\alpha = 1$. The dashed graph of extraction in Figure 17.5 shows $10(y-1)$, ten times the water use in excess of the recharge, 1. Multiplying by 10 makes the scales of extraction and stock similar.

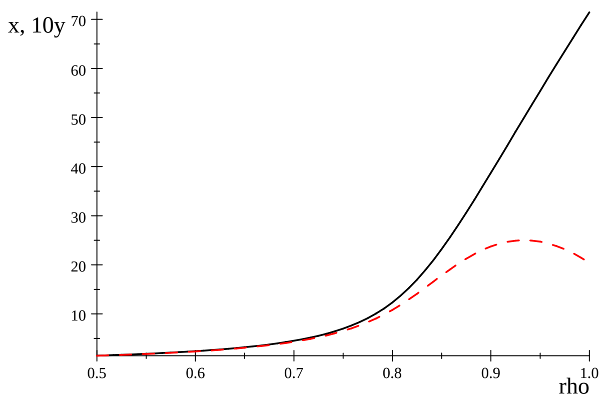

Figure 17.5: The steady state stock (solid) and 10 times the steady state extraction from the aquifer steady state extraction, as a function of ρ

in order to increase the stock, because doing so lowers the cost of extraction. For discount factors below about 0.4, both the the stock and extraction from the aquifer are approximately 0 in the steady state. The steady state single period payoff and rent both increase with the discount factor. For discount factors below about 0.4, the planner cares so little for the future, that the steady state rent and payoff are both negligible. The rent falls to 0 at $\rho = 0$.

We noted in Chapter 5 that with a nonrenewable resource, extraction (and possibly also the stock) eventually approaches 0. In the renewable resource context, we saw in Chapter 15.3 that it might be optimal to drive a stock to extinction if the growth rate is small relative to the discount rate.[6] For a slowly growing renewable resource such as groundwater, the growth rate is very small. Thus, it might be optimal to eventually exhaust an aquifer with a low recharge rate, even if the discount factor is quite large. This caveat reveals a limitation of steady state analysis. The steady state might be insensitive to the discount factor even if the extraction path and welfare are very sensitive to it. By focusing exclusively on the steady state, we might mistakenly conclude that the discount rate in not important to the planning problem. For the same reason, we might mistakenly conclude that the common property and sole owner (= social planner) outcomes are similar, simply because their steady states are similar or the same. The steady

[6] Keep in mind the relation between the discount rate, r and the discount factor, $\rho = \frac{1}{1+r}$. A small discount rate corresponds to a large discount factor.

state may be less interesting than the path that takes us to the steady state, particularly when the growth rate is very small. Exhaustion might occur in 50 years under common property, but in 200 years under the social planner. The fact that the steady states are the same does not imply that welfare is similar under the two trajectories.

> *Box 17.1 Uncertainty.* We consider only the deterministic model. Several sources of uncertainty, e.g. involving inflows, the function F, and the level of x, are important, but including them greatly complicates the model. One insight about the stochastic setting is readily available. We emphasized that in a static setting, efficiency requires arbitrage over competing uses; in a dynamic setting, efficiency requires arbitrage over time. In the setting with uncertainty, efficiency requires "arbitrage over states of nature". Consider the simplest setting in which random inflows might be high or low, $\alpha \in \{\alpha_H, \alpha_L\}$. We speak of the two possible realizations, α_H and α_L, as "states of nature". "Arbitrage over states of nature" means that we adjust the extraction decision depending on α; for example, we might decide to use more water if inflows are high.

17.3.3 A common property game

We show how the outcome changes as we move from full regulation to common property with $n \geq 1$ self-interested farmers. If $n = 1$, we have a sole owner; for large n, the common property problem becomes severe. We use a two-period example to identify the consequence of increasing n, and then return to the dynamic water model.

A two-period game

For the two-period model, suppose that Farmer i obtains the benefit (net of extraction cost) $B\left(y^i\right)$ from extracting y^i in the first period. In the second period, the remaining water, $x_1 = x_0 - \sum_{j=1}^{n} y^j$ will be split equally among the n farmers, and each will have the present value benefit $\rho W\left(\frac{x_1}{n}\right)$. ($B$ and W are concave functions, but otherwise unrestricted.) The social planner wants to maximize the aggregate welfare of all farmers. This planner's objective and first order conditions are

$$\max_{\{y^1, y^2 \ldots y^n\}} \left(\sum_{j=1}^{n} B\left(y^j\right) \right) + \rho n W\left(\frac{x_1}{n}\right) \Rightarrow B'\left(y^i\right) = \rho W'\left(\frac{x_1}{n}\right), \; i = 1, 2 \ldots n. \tag{17.7}$$

Farmer i's objective and first order condition are

$$\max_{y^i} B\left(y^i\right) + \rho W\left(\frac{x_1}{n}\right) \Rightarrow B'\left(y^i\right) = \frac{\rho}{n} W'\left(\frac{x_1}{n}\right). \tag{17.8}$$

The first order conditions for the individual farmer and the planner differ because the farmer weighs the next period marginal benefit, W', by $\frac{\rho}{n}$, whereas the planner weighs the next period marginal benefit by ρ. The farmer knows that if she consumes one more unit of water in the first period, her subsequent allocation will fall by $\frac{1}{n}$. She does not take into account the fact that her additional first period consumption reduces the subsequent allocation of the remaining $n-1$ farmers. The planner, in contrast, takes into account that by giving Farmer i an additional unit of water in the first period, all farmers' subsequent allocation falls by $\frac{1}{n}$ The marginal loss to all of these farmers is $nW'\left(\frac{x_1}{n}\right)\frac{1}{n} = W'\left(\frac{x_1}{n}\right)$.

This two-period example shows that in moving from the social planner to common property with n farmers, it is as if the discount factor falls from ρ to $\frac{\rho}{n}$. The discount factor does not literally change: it is constant at ρ. However, an agent attaches less value to conserving a resource stock when she knows that other people will obtain some of the benefit of her conservation.

The dynamic game

The details are more complicated in the multiperiod setting, but the same basic idea holds. Here, we need to compare the Euler equation under the social planner with the Euler equation for an individual farmer in a noncooperative Nash equilibrium. This comparison identifies the two externalities that lead to excessive extraction in the common property game: the "cost externality" and the "scarcity externality". Both the cost and the scarcity externalities lead to higher extraction and lower welfare under common property, compared to under the sole owner or social planner. In a Nash equilibrium, the actions of individual farmers are individually efficient, but collectively inefficient.

The cost externality arises because in making her own current extraction decision, the individual farmer does not take into account that the lower future stock caused by her extra unit of extraction at t, raises future extraction costs for all of her neighbors. She only takes into account the effect of her extraction on her own future costs.

The scarcity externality arises because the individual farmer understands that her neighbors will likely condition their future extraction decisions on the future stock. A lower future stock increases neighbors' extraction costs and also makes the resource more scarce. Both of these features lower the neighbors' incentive to extract. Therefore, an individual farmer understands that by extracting an extra unit today, her neighbors will (likely) reduce their future extraction. Under the sole owner (or social planner), an additional unit of extraction today takes that unit away from the owner/planner in the future. In contrast, under common property, extraction of an additional unit by an individual farmer today, takes only part of that unit away from that farmer in the future; it also takes part of that unit away from her neighbors. The neighbors' losses are external to the individual farmer's decision problem. Appendix K provides details of this game.

17.4 External trade under common property

Our assumption that farmers within the aquifer are identical eliminates any incentive that they have to trade amongst themselves. However, it leaves the possibility that they might want to trade with water users outside their aquifer. If that trade is allowed, it would occur whenever rent differs across the two regions. A water market that allows northern states to ship water to Texas would increase $V'(y)$ from the exporting regions; the water extracted in those regions now has a more profitable use: exports. Trade therefore would tend to increase pumping from parts of the Ogallala aquifer where stocks remain high. In the absence of other market failures, this interstate market increases social surplus, because it transfers water from a region where the value of marginal product of water is low, to one where it is high. However, the common property problem means that pumping is already too high. The creation of the interstate water market reduces one distortion (the spatial difference in the value of marginal productivity of water) at the cost of worsening another distortion (excessive pumping from the aquifer). If the second distortion is more serious, as is likely the case with the Ogallala aquifer, the water market reduces welfare. (Recall the Theory of the Second Best, Chapter 9.)

This trade-related concern is even more important in the international context. The Theory of Comparative Advantage explains why trade is mutually beneficial for countries with different relative production costs.[7] This theory assumes a first-best setting, e.g. the absence of common property distortions. Many resource-rich countries, particularly, developing countries, have weak property rights for natural resources. Their resource abundance and weak property rights both contribute to low domestic resource prices (e.g., cheap water or forest products). When they open up to international trade, their low domestic prices make them an attractive source for foreign buyers, resulting in exports of natural resources or of commodities that use natural resources for production.

To the extent that these countries' low domestic resource prices derive from resource abundance, they have a "real" comparative advantage in the resource sector, and international trade tends to increase their welfare. However, to the extent that their low domestic price derives from weak property rights (leading to excessive extraction) they have an "apparent" but not a real comparative advantage. In that situation, trade exacerbates a market failure and possibly reduces their welfare. The market does not care about, and cannot distinguish between, real and apparent comparative advantage.

Trade in mammals for which there are weak or nonexistent property rights is likely to harm resource stocks. Examples include seals, beaver, the Arctic Bowhead whale, buffalo, elephants, and rhinos. The trade-resource nexus is equally or more important for forests, fish stocks, and water supplies. There, identifying the impact of trade is particularly difficult, because in many cases trade liberalization and migration (or population increases) occur during the same period. Lack of data makes it hard to separate the effects, on resource extraction, of these confounding factors. Consequently, most of the empirical literature on natural resources and trade relies on case studies, not econometric methods. It would be impractical to obtain a large random sample of cases. Instead, cases are chosen because they are likely to exhibit an important trade-resource connection. The fact that many of these studies finds such a connection, does not imply that it exists in general. This limitation occurs because of the non-random selection of cases, and is endemic

[7] This theory states that trade increases *aggregate* welfare in a country. Trade typically harms some agents in some countries.

to the case study approach.

It is nevertheless worth noting that many case studies find that trade aggravates resource degradation. In some situations, e.g. in Argentina and Senegal, trade and investment liberalization contributed directly to overharvesting of fish stocks. Here, an additional distortion, EU subsidies to EU fleets, compounded the problem of weak domestic property rights. Other examples show why there is not a simple relation between trade and resource use. An EU policy to stimulate livestock production in Ile de la Reunion led to a temporary surge in maize exports from Madagascar, accelerating deforestation; however, previous import *restrictions* in Madagascar, aimed at increasing domestic production of food, led to even greater deforestation. In regions of China and Vietnam, shrimp farming for the export market contributed to the decline of mangroves. EU biofuel policy contributed to deforestation (to develop palm oil plantations) in Southeast Asia, eliciting calls for EU policy changes and subsequent complaints of unfair practices to the WTO, by palm oil producers. Trade has complicated effects on natural resources, sometimes benefitting and sometimes harming them. All of these examples involve developing countries. Rich countries face similar, but less pronounced problems. For example, Canada has restricted water exports to the US out of concern that the increased demand would harm Canadian stocks.

Institutions typically adjust more slowly than markets. Extraction rules under common property might have adapted, over a long period of time, to a particular market regime. The rules may be adequate to protect a natural resource, even without formal property rights, when demand is small and relatively constant and local societies are stable. The development-induced migration or the higher demand resulting from trade liberalization might overwhelm existing institutions.

Trade can change incentives to protect natural resources, eventually altering common property rules, or leading to government regulation, or to the creation of property rights. In our water model, trade increases the value of water to landowners above the aquifer, increasing both the incentive to pump it, and also increasing the incentive to protect the aquifer as a means of generating future sales. Either of these forces may dominate; the trade might make the resource so valuable that the property owners (or the government) begin to protect the resource, moving away from common property towards socially optimal extraction. These changes require an intentional political response; they do not arise from the magic of the self-governing market.

17.5 Summary and discussion

Water, like many other resources, is often inefficiently priced. We consume water as a "bundle", consisting of the physical commodity and its location in time and space. If the user price of water includes only the cost of provision (pumping and transportation), and ignores water rents, users pay for only part of the bundle; the resource price is effectively set to zero. (Rent *is* the resource price.) This under-pricing leads to excessive consumption. In many places, users do not pay the full infrastructure costs of provision, associated with building dams or aqueducts. There, the pricing inefficiency is even more severe.

Often inefficient pricing arises from incomplete property rights. People might have the right to a particular flow of a resource, but not be allowed to trade it. This prohibition prevents the resource from being used where its marginal value is highest. In many places, land ownership gives people the right to pump from a common property aquifer, or use some other stock of water, leading to excessively fast extraction. Here, resource use is not arbitraged over time, leading to a dynamic loss in efficiency. The common property problem arises because property rights to land give people access to groundwater, but not ownership of it. The fact that water stocks migrate across the aquifer would make it difficult to assign and enforce such property rights, even if there was the political will to create the rights. Scarcity and future pumping costs, the two sources of rent, are also the two important sources of externality. A farmer's increased pumping raises her neighbors' future pumping costs, and also leaves less in the aquifer for them to use.

The Theory of Second Best is important in water economics, as in other fields of resource economics. Market failures that appear incidental to the problem at hand, might make reform more urgent, or might militate against reform. Crop subsidies create inefficiencies, attracting factors of production (land, labor, water) to the subsidized crop and away from more efficient uses. A water price below the efficient level (a water subsidy) compounds the distortion created by the output subsidy: both attract inputs to the subsidized sector. The water subsidy creates an additional distortion, encouraging the use of water at the expense of other inputs. The crop subsidies tend to magnify the welfare cost of the water distortion (Chapter 9.5).

This chapter illustrates the broader relevance of the tools and the policy questions discussed earlier in the book. To emphasize this generality, we close by noting that the insights obtained from studying common property dynamic resource problems with fish and water, can be applied to forests, the atmosphere, and to other resources. Forests in developed countries are privately or publicly owned. In contrast, forests in many developing countries are *de facto* common property. Weak institutions in parts of Indonesia and Brazil (as elsewhere) enable people to clear forests for their private benefit. Even where property rights are strong, owners do not internalize all of the benefits created by the stock of forests, e.g. biodiversity and carbon sequestration. In this case, private land-clearing and timber harvesting decisions are not socially optimal. (Chapter 15.2.2 contains a related example in the fishery context.) Under common property, extraction decisions have even less regard for these externalities.

The market failure is worse for the atmosphere, for which there are no property rights. Individuals have no (selfish) incentive to restrict their greenhouse gas emissions, leading to dangerous buildup of greenhouse gas stocks. It is not even sensible to think of assigning individual property rights to the atmosphere. Internalization of the externality requires regulation.

All of these resources (fish, water, forests, the atmosphere) involve stock variables. Our collective actions change these stocks. For all of these resources, there are weak and sometimes non-existent property rights, and important externalities. Optimal management of these resources under the fiction of a social planner creates a benchmark. By comparing this benchmark to the common property or open access outcomes, we learn something about the policies or institutional changes needed to induce society to use resources more efficiently.

17.6 Terms, study questions, and exercises

Terms and concepts

Aquifer, consumptive and nonconsumptive uses, water as a "bundle", prior appropriation, block rates, dynamic strategic substitutes, Theory of Comparative Advantage

Study questions

Exercises

1. Suppose that consumer demand is $10 - q$ and the municipality's constant average cost of providing water is C. (a) If the municipality prices water at its constant cost, how much water is consumed, and what is the level of consumer surplus (both functions of C). (b) Suppose that optimal management, rent is R, so the social cost of providing water is $C + R$. What is the optimal level of consumption (a function of C and R). (c) If the municipality can charge only a single price, what price must it charge in order to induce consumers to buy the optimal amount of water? What is the resulting level of consumer surplus. (d) Now suppose that the municipality can charge block rates, and it chooses these rates in order to induce the optimal level of consumption, and also to break even. What is the level of consumer surplus in this situation?

2. Use the perturbation method to derive the social planner's Euler equation 17.4.

3. Using the equilibrium conditions in the two period common property game in Chapter 17.3.3, show that the symmetric Nash equilibrium approaches the open access equilibrium as $n \to \infty$.

4. Using the Euler equation in the dynamic common property game in Chapter 17.3.3, show that the symmetric (Markov perfect) Nash equilibrium approaches the open access equilibrium as $n \to \infty$.

Sources

The Intelligence Community Assessment (2012) provides the quoted statistics on water availability and use, and describes likely consequences of future water problems.

Schoelgold and Zilberman (2007) provide a recent survey of water economics.

R Howitt (1994) provides an empirical analysis of water market institutions.

Baerenklau et al (2015) provide empirical results on the effect of water block pricing in Riverside California.

Schwartzmay (2015) compares water use in Santa Fe, New Mexico, and Fresno, California.

Chong and Sunding (2006) review water markets.

Kassler (2015) reports the California Supreme Court ruling involving tiered water rates.

Nair (2014) describes the causes of the drought in Isfahan, Iran.

Acciavatti (2015) describes the water crisis Uttar Pradesh, India.

Papers in the 2014 Special Collection in Water Resources Research, edited by Krause et al. (2014), study the dynamics of the aquifer-surface water interface.

The USDA Natural Resources Conservation Service website (http://www.nrcs.usda.gov/) describes the Ogallala Aquifer and the problems it faces.

J Braxton (2009) provides an overview of the Ogallala Aquifer.

McGuire (2007) summarizes changes in water levels in the Ogallala Aquifer.

Stewarda et al. (2013) projects water availability over the Ogallalal Aquifer for the next 100 years.

Galbraith (2012) reports on the dispute over the Texas High Plains Underground Water Conservation District rule to restrict groundwater pumping.

Reisner (1986), an engaging history of the water in western states of the U.S., describes the water extraction plans promoted by commissioners of state water conservation boards in the early 1980s.

Dillon (2014) describes the attempt to form farmer-directed conservation groups in Kansas.

Chichilniski (1994) is is among the first paper discussing the role of trade when resource-rich countries have weak property rights.

Copeland and Taylor (2009) provides a model of endogenous property rights under trade; these authors (jointly with J Brander) have made many contributions to this literature.

The summary in Chapter 17.4 is adapted from Karp and Rezai (2015), which discusses endogenous property rights in an overlapping generations framework.

Bulte and Barbier (2011) review applications of the Theory of the Second Best in the trade and resources setting.

Chapter 18

Sustainability

Objectives

- Understand the economic definition and measurement of sustainability.

Skills

- Know the meaning of strong and weak sustainability.
- Understand the meaning and the application of the Hartwick Rule.
- Understand the use of "green national accounts" and other indices of sustainability.

Sustainable development "meets the needs of the present without compromising the ability of future generations to meet their own needs" (*Our Common Future*, 1987). People born in the future are not responsible for, and cannot insure themselves against, our actions. The view that self-interested actions are unethical if they harm people who are blameless, and who cannot protect themselves against those actions, provides a moral foundation for the sustainability criterion.[1] Although the idea of sustainability is straightforward, its measurement is not. A path is sustainable if the stocks of productive assets that we leave our successors are, *in their totality*, at least

[1] "Brute luck" is the outcome of an involuntary and uninsurable lottery; "luck egalitarians" consider it morally wrong to disadvantage others as a consequence of brute luck. One's date of birth is a matter of brute luck, so luck egalitarians consider actions that harm people born in the future unethical.

as great as the stocks that we inherited. These stocks include produced capital (machines, infrastructure), knowledge capital, human capital, and natural resources. The natural capital stocks include inputs for which markets exist, such as oil, copper, fish, and timber, and stocks for which markets are limited or absent, such as biodiversity and a resilient climate.

It is hard to determine whether we are on a sustainable path because it is difficult to measure the changes in these stocks and it is hard to price these changes. Although both of these impediments are important, the second is probably the most serious. Without prices, we cannot evaluate the change in social wealth. A person who inherits money, land, and art, and bequeaths an equal or greater quantity (or value) of each to the next generation, has clearly left more to the future than they inherited.[2] However, if some components are greater and others lower, we need the prices of land and of art to know whether the bequest exceeds the inheritance.

We have observed rising living standards during most of the past two centuries. Rising standards do not imply sustainability, because society may be living off its capital. The stocks of produced capital, knowledge capital, and human capital have risen over the last two centuries, but many stocks of natural capital have fallen (Chapter 1). Society may be in the position of the person who leaves her successors more money but less art.

This chapter provides an introduction and survey of attempts to determine whether our path of development is sustainable. We begin by explaining the concepts of weak and strong sustainability. Using the example above, the criterion of strong sustainability asserts that art cannot be measured in units of money. A bequest satisfies strong sustainability (for this example) if and only if both the bequest of money + (price of land) × (amount of land) and of the stock of art left to the next generation is at least as great as the amount that the first person inherited. If either stock has fallen, the bequest fails the test of strong sustainability.. The criterion of weak sustainability insists that art must have a price (= monetary unit value), even if we cannot measure it precisely. That criterion directs us to estimate this price, and then to determine whether wealth has increased or decreased based on the change in the money-equivalent of the sum of the components of wealth.

[2] It would not make sense to merely count the number of pieces of art. We need some way of assessing the individual values in order to add them up and thus determine the value of the collection. The difficulty of assessing the value of the art collection is analogous to the difficulty of assessing the value of our bequest of natural capital.

We describe and assess the "Hartwick Rule", which prescribes the amount of savings needed to maintain a sustainable path, using the criterion of weak sustainability. The next section describes how economists have attempted to measure changes in wealth. One approach "greens" Gross National Product to take into account changes in a small number of natural resource stocks omitted in standard national accounts. A second approach looks for radically different measures of sustainability, in some cases using the "green" GDP as one of several elements in an index of sustainability; other measures discard GDP and focus exclusively on changes in natural resources.

18.1 Weak and strong sustainability

Objectives and skills

- Know the meaning, and the strengths and limitations of the concepts of strong and weak sustainability.
- Understand the meaning of the Hartwick Rule, and its relation to weak sustainability.
- Understand why a sustainable path exists only if the resource is "not too important" to production.

Strong sustainability requires that capital stocks, including stocks of natural capital, not fall below current levels. Weak sustainability requires that future *utility levels* do not fall below the current level. These two concepts address different aspects of the sustainability question. The criterion of strong sustainability is conceptually simple, but neglects the possibility of substituting one type of good for another, both in the production process and in consumption. Weak sustainability allows any type of substitution possibility; implementing this criterion requires making specific assumption about substitutability. Both concepts require measure of changes in stocks of natural capital. We have estimates of stocks of resources for which there are markets. Stocks of other resources, or aggregations of those stocks, such as biodiversity, are important, but extremely hard to measure.

The concept of strong sustainability not only ignores the possibility of substitution, but in the case of nonrenewable resources, it sets an impossible standard. We cannot, for example, consume oil while also keeping the stock of oil from falling. Why should we even want to do so? Higher stocks of produced capital, technology, and human capital, which require current oil consumption, might reduce and eventually eliminate our dependence on oil. In the case of renewable resources, strong sustainability might be feasible, but because of opportunities for substitution, not be a sensible objective. Farmland is a renewable resource; we can use it without depleting its stock. If our concern is with agricultural production, not farmland *per se*, then technological improvements that increase yield make it possible to increase production and also convert some farmland to other uses. That conversion violates the strong sustainability criterion, but might be in society's interest.

The concept of weak sustainability uses production functions and a utility function to aggregate all goods and services into something we call utility. Utility can depend on material goods, such as TV sets, services created by natural resources (e.g. hiking and fishing opportunities), and even on the existence of (rather than the use of) natural resources. By choosing the production and utility functions, we can impose any degree of substitutability; these function choices are often driven by mathematical convenience.

An additional dollar of income contributes to people's utility by enabling them to buy more goods. The additional utility produced by one more dollar of income is the marginal utility of income. The ratio of the marginal utility of the resource based good to the marginal utility of income provides a measure of people's willingness to substitute between market-based goods and resource-based goods. The units of marginal utility of income are $\frac{\text{utility}}{\text{dollars}}$, and the units of the marginal utility of a resource good are $\frac{\text{utility}}{\text{resource good}}$. The units of the ratio of these two ratios are

$$\frac{\frac{\text{utility}}{\text{resource good}}}{\frac{\text{utility}}{\text{dollars}}} = \frac{\text{dollars}}{\text{resource good}},$$

the dollar equivalent of the marginal utility of the resource good.

Travel cost models and contingent valuation surveys provide ways of estimating this ratio. Travel cost models use data on the amount of time and money people spend in reaching places where they can (for example) fish or hike, to estimate the implicit price they pay for those recreational opportunities. Those implicit prices, together with data on how often people go fishing or hiking, can be used to construct the value to them, and thus to society, of a marginal change in recreational opportunities. These models are difficult to estimate and they involve assumptions that cannot be tested. In addition, natural capital typically has value beyond the recreational services it provides, so at best these models can measure one component of the value of natural capital. Contingent valuation surveys ask people how much they would be willing to spend (e.g. in the form of higher taxes) to achieve a particular environmental outcome. These methods provide information about people's willingness to substitute income for (typically small) changes in natural resources. They provide limited information about our willingness to make non-marginal substitutions across goods/services associated with natural capital versus produced capital.

Some sustainability measures use a hybrid of the weak and strong concepts. For example, the Ecological Footprint (EF) measures the number of hectares, of average productivity, it takes to sustain a population of a given size at current levels of consumption. In aggregating all natural capital into a "hectare equivalent", the measure implicitly assumes perfect substitutability across the different types of natural capital; in this respect, EF uses a concept similar to weak sustainability. However, the measure does not include produced capital or technology, implicitly assuming zero substitutability between natural and produced capital; in this respect, EF uses a concept similar to strong sustainability.

18.1.1 Weak sustainability

Weak sustainability accommodates nonrenewable resources, where stocks fall while extraction is positive, and also takes into account substitutability across goods and services. We address two questions about (weak) sustainability: (i) What investment policy leads to a sustainable path? (ii) When is a sustainable path feasible? Our model assumes constant population and a single "composite commodity", i.e. a single consumption good. The composite commodity assumption means that we cannot discuss substitution across goods in consumption, but it enables us to address substitution across inputs to production. The two inputs are the stock of man-made capital, $K(t)$ and a resource flow (e.g. oil, or more generally, energy produced using nonrenewable resources), $E(t)$. These inputs produce the composite commodity, $Y(t) = F(K(t), E(t))$ under the following assumptions:

(i) $F(\cdot)$ is constant returns to scale in K, L: doubling both inputs doubles output.

(ii) Both inputs are necessary to production: output equals 0 if either input is 0.

(iii) Man-made capital, K, does not depreciate.

(iv) The constant average cost of extracting the resource is c.

By choice of units, we can set the price of the composite commodity equal to 1, making it possible to interpret Y as both the physical amount of the commodity, and the value of the commodity (= income). Therefore, $\frac{\partial F}{\partial K}$ represents both the marginal product of capital and the *value* of marginal product of capital. We denote $p(t)$ as the price of a unit of energy, and $r(t)$ as the rental rate for capital. The competitive equilibrium conditions require that the price of an input equals its value of marginal product:

$$\frac{\partial F(K(t), E(t))}{\partial K} = r(t) \text{ and } \frac{\partial F(K(t), E(t))}{\partial E} = p(t). \tag{18.1}$$

The assumption of constant extraction costs implies that the resource rent at time t equals $p(t) - c$. The Hotelling rule for competitive extraction of a nonrenewable resource with constant extraction costs states that rent rises at the rate of interest, $r(t)$ (cf. equation 5.9). In the continuous time setting adopted here, this rule is

$$\text{Hotelling Rule: } r(t) = \frac{\frac{d(p(t)-c)}{dt}}{p(t) - c}. \tag{18.2}$$

The stock of capital (K) and the resource stock (x) evolve according to:

$$\text{Definitions: } \frac{dK}{dt} = I \text{ and } \frac{dx}{dt} = -E.$$

One unit of investment, $I(t)$, adds one unit to man-made capital. One unit of energy, E, reduces the resource stock by one unit.

18.1.2 The Hartwick Rule

A sustainable consumption path requires that utility remain constant; because utility depends only on consumption (in this simple model), a sustainable consumption path requires a constant level of consumption. If indeed a sustainable consumption path exists, the "Hartwick Rule" states that this path requires that society invest (rather than consume), at each point in time, the resource rents. Using $\frac{dK}{dt} = I$ and the definition of resource rent, the rule is (Appendix L.1):

$$\text{Hartwick Rule: } I(t) = (p(t) - c)\, E(t)\,. \tag{18.3}$$

The Hartwick Rule has an intuitive explanation. If there were a single factor of production, e.g. a single capital stock, then it would be obvious that maintaining a constant level of consumption (and thus utility), requires maintaining a constant capital stock. Our model, however, has two factors of production, capital and the resource input. Moreover, it is not feasible to maintain a constant positive level of resource extraction, because doing so would eventually exhaust the resource stock. After exhaustion occurs, extraction drops to 0, at which time output and consumption also equal 0. However, by building up the stock of man-made capital, K, society may be able to decrease resource use over time, approaching (but never reaching) zero resource use. The resource stock falls, but it is not exhausted in finite time. Under the assumption that it is possible to achieve this delicate balancing act, the Hartwick Rule explains how it is done: by investing resource rents in man-made capital.

18.1.3 Existence of a sustainable path

When is it possible to maintain a constant consumption trajectory? For the special case where the production function is Cobb Douglas, $F(K, E) = K^{1-\alpha}E^{\alpha}$, with $0 < \alpha < 1$, the answer can be illustrated geometrically. The parameter α equals the revenue of the resource sector, pE, as a share of the value of output, $F(K, E)$: $\alpha = \frac{pE}{F}$. A smaller value of α means that the resource sector contributes a smaller fraction of value added to the economy. A sustainable trajectory is feasible if and only if $\alpha < 0.5$; this inequality states that the resource is not "too important" in production.

With Cobb Douglas technology and the Hartwick Rule, consumption equals the fraction $1 - \alpha$ of output; remaining income is invested or pays for extraction. Here, constant consumption requires constant income. This fact means that the question can be rephrased as "When is it feasible to maintain constant income?" To answer this question, pick an arbitrary positive level of income, Y. With Cobb Douglas technology, we can rewrite the relation $Y = K^{1-\alpha}E^{\alpha}$ to express E as a function of Y, K, and α: $E = (YK^{\alpha-1})^{\frac{1}{\alpha}}$.

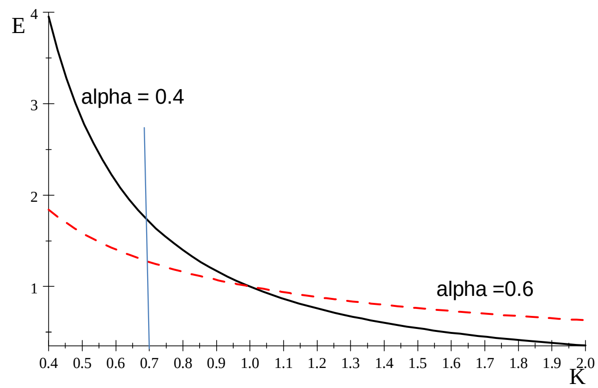

Figure 18.1: Isoquants for two values of α, with $Y = 1$. Given initial stock $K = 0.7$, the area under each isoquant, to the right of 0.7, equals the size of the initial resource stock needed to maintain constant output Y when society follows the Hartwick Rule. Area = 2.4 for $\alpha = 0.4$ and area is infinite for $\alpha \geq 0.5$.

The curve showing E as a function of K, for a particular value of Y, is an isoquant: the combination of E and K needed to produce Y.

Figure 18.1 shows the graphs of two isoquants, corresponding to $\alpha = 0.4$ and $\alpha = 0.6$, for $Y = 1$.[3] The $\alpha = 0.4$ isoquant lies above the $\alpha = 0.6$ isoquant for small levels of K, but crosses it and falls more steeply toward 0 as K increases. This relation shows that for large capital stocks, the technology corresponding to $\alpha = 0.4$ requires less more of the resource input (compared to the technology with $\alpha = 0.6$) in order to produce $Y = 1$.

We can check whether it is feasible to maintain a constant stream of income. With constant income, savings remain positive, so the capital stock continues to grow. Thus, the capital stock becomes infinitely large over time, as resource use (along with the resource stock) falls asymptotically to 0. Given an initial value of the capital stock, constant output requires that the production point "slide down the isoquant". Over time, with increasing capital stock, resource use falls, but Figure 18.1 shows that it falls much faster the smaller is α.

[3] Setting $Y = 1$ is not important to this discussion. With constant returns to scale, the isoquant for any positive value of Y merely scales up or down the isoquant corresponding to $Y = 1$.

Suppose that the initial capital stock is $K = 0.7$ (the location of the vertical line in Figure 18.1). The area under the isoquant, from $K = 0.7$ to $K = \infty$, equals the initial resource stock needed to maintain a constant level $Y = 1$ of output (and therefore a constant level of consumption) when society follows the Hartwick Rule. For $\alpha = 0.4$ and $K = 0.7$, society needs to begin with 2.4 units of the resource in order to sustain a constant $Y = 1$. For any $\alpha \geq 0.5$, society would need an infinitely large initial resource stock in order to maintain the constant level of output. Therefore, the constant output trajectory $Y = 1$ (and indeed, *any* constant trajectory with $Y > 0$) is not feasible when $\alpha \geq 0.5$. Figure 18.1, makes it plausible that the area under the isoquant is much smaller under $\alpha = 0.4$ compared to $\alpha = 0.6$, simply because the curve falls so much more quickly if $\alpha = 0.4$.

If $\alpha \geq 0.5$, it is not feasible to maintain in perpetuity any positive constant level of output (or consumption). In this case, there is no sustainable plan. If $\alpha < 0.5$ and the initial capital stock is positive, it is possible to support a positive sustainable consumption path, one that depends on the initial stocks of capital and the resource, and on α.

18.1.4 Adjustments to the Hartwick Rule

We described the Hartwick Rule in the simplest setting, with a single man-made stock of capital and a single stock of nonrenewable natural capital, and restrictive assumptions about technology. The model shows that in some circumstances, by investing natural resource rent into man-made capital, i.e., transforming natural capital into man-made capital, it is possible for society to sustain a constant level of consumption. In these circumstances, a society that follows the Hartwick Rule can gradually use up resource, without harming future generations. In other circumstances, it is not possible to maintain forever any positive consumption level; resource constraints imply that consumption must eventually fall to 0.

This model can be generalized by including depreciation of man-made capital, renewable resources (fish, not just oil), and the inclusion of many stocks of both man-made and natural capital. The generalization to many stocks accommodates knowledge as well as physical capital. By investing in knowledge capital (education, research and development), society changes the technology, likely relaxing the resource constraint. Empirically, a higher stage of economic development (greater wealth) is associated with a decrease in the number of units of energy per unit of output; this negative correlation between wealth and energy intensity is particularly strong for individual countries, as they develop (Table 18.1).

Country	1971	1990	2000	2005	2010
United States	0.41	0.27	0.23	0.21	0.19
Germany	0.29	0.20	0.16	0.15	0.14
Japan	0.22	0.15	0.16	0.15	0.14
Korea	0.21	0.22	0.23	0.21	0.20
Brazil	0.17	0.14	0.15	0.15	0.15
China	0.88	0.47	0.22	0.21	0.19
World	0.32	0.26	0.22	0.21	0.19

Table 18.1: Total primary energy supply per unit GDP
Tones of oil equivalent per thousand 2000 US dollars using PPP[4] (OECD Factbook 2011- 12)

Endogenous technical change does not fundamentally alter the conclusions discussed above, provided that production is constant returns to scale. Increasing returns to scale at level of the economy can increase the possibility of a sustainable *level* of utility (as considered above) and even of sustainable *growth.* For example, if the production function is $Y = NK^{1-\alpha}E^{\alpha}$, where N denotes knowledge capital, then doubling all inputs, N, K, E, leads to a four-fold increase in output. In this setting, sustainable consumption may be possible even if $\alpha > 0.5$; whether it is actually possible depends on the cost of increasing knowledge capital.

Even if sustainability is feasible, there is no presumption that society actually follows a sustainable path. Current generations might want to consume some of the resource rent, violating the Hartwick Rule and leading to decreased consumption. Even if consumption remains constant for a period of time, this level of consumption may be unsustainable if it includes a portion of the resource rents or if a sustainable path is not feasible.

The model above involves a “closed economy”, one where there is no trade. It is thus appropriate for describing the aggregate world economy: our world currently cannot trade with any other world. The model is not designed to describe an open economy, particularly one that is “small” (i.e. a country whose trade does not affect the prices at which it exchanges goods with other countries). To see how trade changes matters, consider a country

[4] PPP is purchasing power parity, a method of converting foreign currency to US dollars based on prices of a reference bundle of commodities, instead of the official exchange rate.

that has no resource stock, and therefore must import E. Suppose that this importer has the production function used above, $Y = K^{1-\alpha}E^{\alpha}$. Because this importer has no resource stock, it earns zero rent. In this case, the Hartwick Rule says "invest nothing". If the import price of the resource were constant, this rule would indeed lead to a sustainable consumption path. The importer buys a constant amount of oil at a constant price, exports the fraction α of its income = output in order to pay for resource imports, and consumes the remaining fraction.

If the import price increases over time (as in the Hotelling model), the importer must increase its capital stock in order to maintain a constant level of consumption. The increase in capital stock requires investment, lowering consumption. Because the importer with no resource stock has zero rent regardless of whether it faces increasing import prices, the Hartwick Rule instructs it to invest nothing. But if it invests nothing, its consumption path falls over time, as the import price increases. In this open economy that faces increasing import prices, the Hartwick Rule leads to a decreasing, not a sustainable consumption path.

The case of a resource exporter is the mirror image. For example, Nigeria obtains most of its foreign revenue from oil exports and has little influence on the world price of oil. Increases in oil prices increase the value of its resource stock, thereby increasing Nigeria's wealth, making it possible for Nigeria to maintain a sustainable consumption path while also consuming some of its current rent. In this case, the Hartwick Rule instructs the economy to consume too little (for the purpose of maintaining constant consumption). The Hartwick Rule pertains to a closed economy, so it requires modification if used for an open economy.

18.2 Welfare measures

We would like to know whether future generations will be better or worse off than current generations. This comparison requires a measure of welfare. We first describe how standard national accounts can be modified to take into account resource depletion and other changes that affect sustainability. We then consider alternative welfare measures.

18.2.1 Greening the national accounts

Economists use Gross Domestic Product (GDP) or the closely related Gross National Income (GNI) as measures of national wellbeing. (Box 18.1) To simplify the discussion, we ignore foreign remittances and assume that trade is balanced (value of imports = value of exports); here, GDP and GNI are the same. In the simplest setting, net output (defined as output after replacing depreciated capital) depends only on the stock of capital, K: net output equals $F(K)$. Output can be used either for consumption, C, or net investment, I, so $GDP = F(K) = C + I$.

> *Box 18.1 Measures of income.* GDP and GNI measure economic activity within a country's borders, and GNP measures economic activity of the country's residents.
> GDP = consumption + investment + government spending + exports - imports.
> GNP = GDP + net income receipts from assets abroad minus income of foreign nationals in the country.
> GNI = GDP + payments into the country of foreign nationals' interest and dividend receipts, minus similar payments out of the country.
> Example 1: The output of an American owned factory in China contributes to China's GDP. Profits from this factory that are repatriated to the US reduce China's GNP and increase US GNP. Example 2: Profits that a foreign national living in the US earns outside the US and brings into the US, contributes to US GNI but not to US GDP.

Estimated adjusted net savings and population growth are inputs to estimates of changes in per capita wealth. The World Bank estimates that almost half of the world's countries have falling per capita wealth. Amongst 24 low income countries, and 32 Sub-Sahara African (SSA) countries, almost 90% having falling per capita wealth. (The annual increase in population in SSA is about 2.7%, much lower than adjusted net savings.)

Wealth is a stock variable and income is a flow variable. Wealth and income are highly correlated, but they are not the same thing. A rich person's primary source of income is often the return on their invested wealth. Their income might fluctuate, even though their wealth remains high. Wealth provides a better indication (compared to income) of a person's future consumption possibilities. GDP is better measure of income than wealth, but in some cases it is closely related to wealth. Suppose that the consumption trajectory, $C(t)$, is the equilibrium to a competitive economy, and that the discount rate for consumption is the constant r. As above, production depends only on capital, and output can be either consumed or invested. We *define* society's wealth, or welfare, as the present discounted value of the stream of future consumption, $W = \int_{t=0}^{\infty} e^{-rt} C(t)\, dt$. It can be shown that

$$rW(t) = C(t) + I(t) \quad (= GDP(t)). \tag{18.4}$$

It is as if society, with wealth W, can invest in an asset that pays the return r; society uses the return on the asset, rW, for the purpose of consumption and net investment. Equation 18.4 helps to explain why GDP is a proxy for wealth: in some cases they are related by a factor of proportionality, r.

The introduction of nonrenewable resources, or other stock variables, does not change the basic idea, provided that these resources are priced efficiently. Efficient pricing requires secure property rights and well-functioning markets for the resource. In order to incorporate natural resources and maintain a simple model, suppose that production depends on capital and on a single nonrenewable resource, e.g. oil; suppose also that oil can be extracted costlessly. When extraction of the resource is E, output is $F(K, E) = C + I$. In a competitive equilibrium, we know from Chapter 5 that the price of this resource, denoted $p(t)$, rises at the rate of interest; moreover, the price equals the resource rent, or the shadow value of the resource. In this setting, with secure property rights and efficient markets, the resource price is an accurate measure of the value (both to the resource owner and to society) of an additional unit of the resource stock. The reduction in society's stock of capital, due to the extraction of a unit of the resource is $p(t)\,E(t)$. The proper measure of "adjusted" GDP is therefore $C(t) + I(t) - p(t)\,E(t)$, and the measure of wealth becomes

$$rW^*(t) = C(t) + I(t) - p(t)\,E(t). \tag{18.5}$$

The problem arises when the resource price does not accurately reflect the resource's social opportunity cost, as occurs under open access or common property (Chapters 14 and 17). The resource price might be zero, although its social opportunity cost is positive. The atmosphere has limited ability to assimilate CO_2 without leading to costly climate change, so the social cost of carbon is positive; but in most countries the price of carbon is zero. A resource-intensive economy may be generating a high GDP by using up its resources. If we observe $GDP(t)$ and we ignore the resource use (perhaps because it is unpriced) our measure of the society's wealth, using equation 18.4, is $W = \frac{GDP}{r}$. Taking the natural resource into account, our measure of wealth (using equation 18.5), is $\frac{GDP-pE}{r}$. The percent correction, due to properly taking into account resource use, is

$$\frac{\frac{C(t)+I(t)}{r} - \frac{C(t)+I(t)-p(t)E(t)}{r}}{\frac{C(t)+I(t)}{r}}100 = \frac{p(t)E(t)}{GDP(t)}100 \tag{18.6}$$

The social value of resource use $(p(t)E(t))$, as a percent of GDP, gives a measure of the welfare reduction due to resource use. For a resource-based economy, failure to adjust GDP to account for resource depletion may significantly overstate wealth.

The discussion above generalizes to the case of multiple stocks, including renewable and nonrenewable resources and other productive assets such as human capital. Some of these stocks might be growing over time; for example, education or health care can improve the stock of human capital, and conservation efforts can lead to higher stocks of renewable resources. The correction to GDP needed to accurately measure society's wealth, could be positive or negative, depending on how stocks are changing, and on the value of these changes.

Research during the past quarter century has attempted to "green the national accounts" by including the value of changes in productive assets that are not already incorporated into GDP. Estimating GDP requires estimating the value of production of society's goods and services, a daunting measurement problem, but one that has been studied and refined over many decades. Measuring the correction required by changing resource stocks is a harder problem. Researchers have to decide which stocks to include in the correction, then attempt to estimate the reduction in the stock and finally to attribute a price to this stock. For resources with well-functioning markets, such as oil and forestry products, the market price can be used to value the change in stock. However, the correction is also important where property rights and markets are weak or non-existent, requiring researchers to estimate (often with little data) the prices used in the correction.

Table 18.2 shows World Bank estimates of savings, and adjusted savings, for the world and for different regions. After accounting for depreciation ("consumption of capital"), the World Bank estimates that the world saves about $24.5 - 13.6 = 10.9\%$ of GNI. Educational investments, which increase the stock of human capital, almost exactly offset the reductions due to resource depletion and pollution, so adjusted net savings are also close to 11% for the world. Gross savings rates and consumption of capital in Sub-Sahara Africa (SSA) are close to world levels. Investment in education is slightly lower than world levels, but the correction for resource depletion and pollution damages is much higher, resulting in an estimated adjusted net savings for SSA of 0.9%. Increases in population imply that per-capita wealth is falling in these regions.

	World	EAP[1]	LAC[2]	MENA[3]	SSA[4]
Gross savings	24.5	47.6	19	25.9	26.3
(−) Consumption of fixed capital	13.6	12.0	12.2	9.9	13.0
(+)Educ. Expenditure	4.3	2.1	5.1	4.5	3.4
(−)Energy depletion	2.4	2.7	4.7	12.9	10.3
(−)Mineral depletion	0.6	1.4	1.2	0.5	1.8
(−)Net forest depletion	0.1	0.1	0.4	0.2	1.8
(−)CO_2 damage	0.5	1.0	0.3	0.7	0.6
(−) Particulate emissions damage	0.6	1.6	0.8	0.9	1.2
Adjusted net savings	11.1	30.0	4.5	5.3	0.9

Table 18.2 National accounting aggregates (savings, depletion and degradation). All numbers are percent of Gross National Income. 1 East Asia and Pacific; 2 Latin America and Caribbean; 3 Middle East and North Africa; 4 Sub-Sahara Africa. Source: World Bank, Little Green Data Book 2014.

18.2.2 Alternatives to adjusted national accounts

GDP or GNI include some components that do not belong in a measure of welfare, and exclude some that do belong. Increased construction of prisons and employment of prison guards might stimulate the economy, increasing overall employment and GDP. If this increased activity is the result of stricter laws for minor infractions, and if those laws contribute to social dysfunction, the additional prisons and the guards do not represent an increase in social welfare. GDP measures only market-based transactions. If a couple divorces and one person begins paying for services that were previously unpaid, those payments (if recorded) show up as an increase in GDP. However, the change likely does not represent a real increase in economic output, or welfare. GDP statistics do not reflect inequality, which may reduce social cohesion, lowering welfare. Higher levels of pollution or congestion likely decrease welfare, but because these are typically unpriced, GDP does not capture them. Green national accounts attempt to remedy this omission, but not the others.

A literal interpretation of strong sustainability is impractical, because it would require a long list of stocks, many of which we have no hope of measuring. Even if it were possible to measure the components of this list, it would be too complex to understand, and therefore useless for policy guidance. A useful welfare measure must present information in an intelligible manner. The simplicity of national accounts such as GDP is an important part of their appeal. Politicians routinely use changes in GDP as evidence of their own or their rivals' economic (in)competence. An index combines different pieces of information into a single number. GDP adds up the value of goods and services in an economy. A green national account includes the estimated value of unpriced (or mis-priced) goods and services. Because all of these components are in the same units (e.g. dollars) it is sensible to add them together. For indices that involve non-commensurable components, merely summing the components is arbitrary.

The United Nations (UN) produces the Human Development Index (HDI), a widely used index of welfare that includes measures of health, education, and material wellbeing. The HDI aggregates these three components using their geometric mean (the cube root of the product of the components). A 1% change in any of the components has the same effect on the geometric (but not the arithmetic) mean. The geometric mean also implies less substitutability among the different components, compared to the arithmetic mean. The UN also produces broader indices of well-being that include factors such as measures of inequality, human security, and gender disparity.

Other indices include the Measure of Economic Welfare (MEW), Sustainable Measure of Economic Welfare (SMEW), Index of Sustainable Economic Welfare (ISEW), the Genuine Progress Indicator (GPI), and Ecological Footprint (EF). MEW adds (to standard national accounts) the estimated value of activities that contribute to welfare (e.g. leisure) and subtracts activities that do not (e.g. commuting). SMEW modifies MEW by taking into account changes in wealth. ISEW and GPI deducts other costs, including those related to pollution, the loss of wetlands, and CO_2 damage. Over the past quarter century, GDP has continued to grow, whereas alternatives such as the GPI and the HDI have been flat. Different measures of welfare lead to different conclusions about sustainability.

The EF calculates the amount of "average quality" productive land needed to support a population at current consumption levels. Our species' EF exceeded the earth's carrying capacity by 25% in 2003. There are about 1.8 hectares of average quality land per person available globally; Europeans use about 5 hectares per person, and North Americans use twice that amount. The EF takes into account the forest area needed to absorb carbon emissions. Changes in consumption levels or in production methods could alter our EF; those sorts of changes caused Malthus' predictions to not (yet) occur. Cross country differences in $C0_2$ emissions explain a large part of the cross country differences in EF. Compared to its EF, a country's carbon footprint proves a more easily calculated and communicated measure of its resource use.

18.3 Summary

Economic development is sustainable if it meets current needs without sacrificing the ability of future generations to meet their needs. Attempts to rigorously define and to measure sustainability rely on concepts of weak or strong sustainability. The former recognizes substitutability in production and consumption, and focuses on future utility levels. The later assumes limited substitutability, and focuses on maintaining constant or increasing levels of stocks.

A body of theory studies the decisions needed to achieve weak sustainability. In a simple setting, weak sustainability requires that resource rents be invested in man-made capital. This investment program transforms natural capital into man-made capital, achieving weak sustainability if and only if natural capital is not "too important" in production. This conclusion rests on many assumptions, but is intuitive: a constant or increasing stream of future utility is feasible only if man-made capital provides an adequate substitute for a dwindling supply of natural resources. Even if weak sustainability is feasible, there is no reason to assume that society is on a sustainable trajectory.

Attempts to measure sustainability have followed two principle avenues, closely related to the concepts of weak and strong sustainability. The first begins with the positive relation between wealth (a stock) and GDP (a flow). Standard national accounts (e.g. GDP) measure the value (in dollars, or some other currency) of market-based economic activity. These statistics ignore the value of changes in many natural resources and in other stocks, e.g. in human capital. During the last quarter century, economists have attempted to include the value of these kinds of changes, resulting in "green" national accounts. Recent estimates show that many poor countries are not increasing their stock of man-made plus natural capital fast enough to accommodate population growth. By this measure, these countries appear not to be on a (weakly) sustainable development path.

The second approach to measuring sustainability focuses on resource stocks, not standard economic measures of income. There are many of these measures; some rely on a single number, e.g. the amount of average quality productive land needed to support a population, in perpetuity, at current levels of income. This measure concludes that our development trajectory is not sustainable, because the actual population exceeds the level that can be supported by available land. Other measures create indices that aggregate measures of health, education, material wellbeing, and sometimes other components. These indices attempt to provide welfare measures, without necessarily enquiring whether this level of welfare is sustainable.

The variety of measures of sustainability (or welfare) is testimony to the difficulty of the empirical question. At a sufficiently abstract level, it is easy enough to say what we think sustainability means (even if there is disagreement on this point). However, even under a host of assumptions, it is not easy to reach definitive conclusions about the sustainability of our development path, i.e. whether future generations are likely to be richer or poorer than the current generation. The large number of different sustainability and welfare measures provide alternatives that focus on different aspects of the same general question.

18.4 Terms and study questions

Terms and concepts

Weak and strong sustainability, travel cost models, contingent valuation, composite commodity, constant returns to scale, Cobb Douglas production function, national income accounting identity, Hartwick Rule, GDP, GNP, GNI, Human Development Index, Ecological Footprint, geometric mean.

Study questions

1. Explain the meaning of weak and strong sustainability; discuss some of the advantages and disadvantages of both concepts.

2. State the Hartwick Rule and describe the question to which it provides an answer.

3. Given the Cobb Douglas production function $F(K,K) = K^{1-\alpha}E^{\alpha}$, state the condition under which a sustainable consumption path is feasible, and provide an intuitive justification for this condition.

4. Using the simple model in this chapter, define GDP. When all inputs are correctly priced, what is the relation between GDP and wealth (defined as the present value of the stream of future consumption. Explain the adjustments to GDP that must be made (in order to use GDP as a measure of wealth) when production uses unpriced (or incorrectly priced) natural resources.

5. Describe the adjustments to gross savings that World Bank makes, in order to calculate "adjusted net savings". How do adjusted net savings vary across countries at different income levels? What is the practical significance of this relation?

6. Green National Accounts and the Ecological Footprint are two attempts to shed light on the issue of sustainability. Briefly explain both of these; your explanation should describe the relation between both of these measures and the concepts of weak and strong sustainability.

Sources

The "Brundtland Report" *Our Common future"* (1987) set out a framework for sustainability, relating economic development and environmental protection.

Roemer (2009) discusses the idea of brute luck and the school of luck egalitarians.

Solow (1974b) and Hartwick (1977) introduced the Hartwick Rule.

Asheim et al. (2003) discuss some of the misconceptions that have arisen related to this Rule. Mitra et al. (2013) summarize and extend results on the issue of sustainability in resource markets.

Table 18.1 is taken from the OECD Factbook 2011-2012.

Table 2 is based on the World Bank Little Green Data Book 2014.

Weitzman (1976b) demonstrates the relation between GDP and welfare.

Weitzman (1999) shows the effect on mineral depletion on welfare.

Hamilton (2002) provides estimates of changes in total and per capita wealth.

Hartwick (2011) explains the relation between green national income and green national product.

Wolff et al. (2011) identified systemic errors in the Human Development Index.

Stiglitz et al. (2009) discuss the theory and the practicalities of measuring economic performance and social progress. They describe the various indices used to measure sustainability.

Kubiszewski et al. (2013) compare measures of sustainability.

Chapter 19

Valuing the future: discounting

Objectives

- Understand the role of discounting in evaluating a policy that has consequences over long spans of time.

Skills

- Understand the basics of the model of discounted (expected) utility.
- Know the difference between discounting utility versus consumption, and understand the "tyranny" of discounting.
- Understand how beliefs about future technology and future wealth affect current policy.
- Understand the relation between impatience and discounting, and the difference between intra- and inter-generational transfers.

Many environmental and resource issues, and climate change in particular, involve welfare trade-offs over long spans of time. How much should society be willing to spend today to reduce the risk of future climate damage? Climate scientists' consensus views provide the proper foundation for evaluating climate policy. However, policy-based models require economic assumptions and ethical judgements, along with climate science. Most of these models use discounted utilitarianism. We describe this framework and explain how it affects policy recommendations.

Carbon taxes have been enacted or seriously considered in only a few countries or regions (e.g. Sweden, and for a time, Australia). Despite its lack of political traction, a carbon tax is useful both for describing policies that are more widely used, and for recommending policies that should be used; these are "positive" and "normative" applications, respectively. Many climate policies that are actually used, including cap and trade, green subsidies, and renewable fuel portfolio standards, can be expressed as a "tax equivalent", a tax that would yield (approximately) the same level of emissions reductions, although usually at different economic cost. Most climate policy models express their policy recommendation (a normative statement) by proposing an optimal tax. A higher tax leads to lower carbon emissions, and thus corresponds to a stricter policy.

The optimal carbon tax equals the "Social Cost of Carbon" (SCC), defined as the present discounted value of the stream of additional costs arising from an extra unit of atmospheric carbon. Chapter 2.5 provides a stylized example of the SCC. An estimate of the SCC requires estimates of the effect of current emissions on future climate variables (e.g. temperature and precipitation) and the link between those variables and economic costs. We use discounting to transform this stream of future marginal costs into a single number, the SCC. Every step of this calculation involves assumptions and judgements. This chapter discusses the use of discounting to aggregate costs across different periods. The material helps readers to evaluate discounting assumptions and to understand how they influence model results.

The U.S. Environmental Protection Agency (EPA) uses the SCC in conducting cost benefit analysis for policies that have significant effects on carbon emissions.[1] In calculating adjusted net savings, the World Bank uses the SCC to estimate the cost of increased atmospheric carbon (Table 18.2). Thus, estimates of the optimal carbon tax are important for policy discussions, despite the fact that carbon taxes are rarely used. Economic models produce a wide range of recommendations for the optimal tax (the SCC), from less than \$10 to well over \$100 per ton of carbon. We do not know whether taxes in this range are too low, too high, or about right, but we can understand and evaluate the assumptions that lead to these estimates.

[1] Most estimates of the SCC consider the cost to the world as a whole, not specifically to the U.S., of an additional unit of atmospheric carbon. The EPA, a U.S. agency, uses a global cost of carbon in assessing the cost/benefit ratio of a U.S. policy.

We begin by explaining discounted utilitarianism. It is straightforward to think of utility as an "ordinal" concept; a person may have no difficulty in ranking ("ordering") two consumption bundles, such as a beer and a pizza versus a movie and popcorn. Deciding that they like one bundle twice as much as the other, or more generally assigning a number of "utils" to each bundle, requires a "cardinal" measure. Policy models involving trade-offs across people or across time use a cardinal measure, a utility function that assigns a number of utils to a particular outcome.

A utilitarian evaluates social welfare by adding up the utility of society's members, assigning welfare $u_M + u_J$ to an allocation (e.g. of consumption) that gives Mary and Jiangfeng utility levels u_M and u_J, respectively. "Discounted" utilitarianism often begins with the fiction of a representative agent at each point in time, proceeding as if all currently living people are identical, or their preferences can be aggregated ("added up") to enable a single agent to represent them. The discounted utilitarian evaluates a stream of utility by discounting utility at each point in time and then adding up the discounted utility levels. Accounting for future uncertainty ("stochastics"), including those related to economic growth and to climate change, requires replacing discounted utility with discounted expected utility (DEU).

The use of *discounted* utilitarianism is a major reason that many economic models support only modest climate policy. Discounting depends on technological optimism ("growth") and impatience. Optimism takes several forms, the most important being that technological change and increased accumulation of (man-made) capital will make people in the future richer than those currently alive. Consequently, climate policy should involve only modest expenses, in order to avoid requiring the relatively poor current generations to make sacrifices that benefit relatively rich future generations. In addition, if future inventions will lower the cost of reducing carbon emissions, it makes sense to delay emissions reductions until they become cheaper. Events of the past two centuries support technological optimism, but provide a questionable basis for policy that might have major effects on our species.

Models that build in impatience for future utility tend to promote modest climate policy. People dislike delaying gratification, even abstracting from the uncertainty about whether they will live to enjoy the future. Because individuals appear to be impatient about their own future utility, some modelers assume that the social planner who acts on their behalf should exhibit the same kind of impatience. This view makes no distinction between intra-personal transfers (from a young person to or from her older self) and intergenerational transfers (from someone currently alive to or from a person who will live in the future).

The next section explains the difference between discounting utility and discounting consumption. It shows the "tyranny" of discounting in influencing policy, and it illustrates the role of uncertainty. The following section develops the idea of consumption discounting, explaining how characteristics of preferences and of the economy determine the consumption discount rate in a deterministic setting. We then explain how economists have introduced uncertainty about growth. The final section takes up the role of impatience, and the distinction between intra- and intergenerational transfers.

19.1 Discounting utility or consumption

Objectives and skills

- Understand the difference between discounting utility and discounting consumption.
- Understand the sense in which discounting is "tyrannical", and the interaction between uncertainty and discounting.

Discounting utility is different than discounting consumption, but in either case it can be "tyrannical", inducing people today to (almost) ignore the future. We show how a particular type of uncertainty interacts with discounting. Throughout this discussion, we use a continuous time setting. For example, if the discount rate under annual compounding is 5%, the discount rate under continuous compounding is about 4.9% (Chapter 2.5).

The discount factor makes objects at different points in time comparable. The logic of discounting is the same regardless of whether we apply it to dollars or utility (or anything else), but the interpretation and the numerical value of the discount rate may vary with the context. To keep this distinction in mind, we use different symbols to represent discount rates applied to different objects. In *this chapter only*, ρ denotes the (continuously compounded) discount rate for utility, and r denotes the discount rate for consumption (measured in dollars). For constant values of ρ and r, the utility and the consumption discount factors are:

$$e^{-\rho t} = \left\{ \begin{array}{c} \text{utility discount} \\ \text{factor} \end{array} \right\} = \left\{ \begin{array}{c} \text{number of units of utility a} \\ \text{person will sacrifice} \\ \text{today to obtain one additional} \\ \text{unit of utility at time } t \end{array} \right\}$$

$$e^{-rt} = \left\{ \begin{array}{c} \text{consumption} \\ \text{discount factor} \end{array} \right\} = \left\{ \begin{array}{c} \text{number of units of consumption (\$)} \\ \text{a person will sacrifice today to} \\ \text{obtain one additional unit of} \\ \text{consumption at time } t \end{array} \right\}.$$

A higher discount rate corresponds to a lower discount factor. A lower discount factor means that we value future utility or consumption less. Valuing the future less means that we are willing to sacrifice less today to benefit the future. *A higher discount rate leads to a lower recommended carbon tax.*

19.1.1 The tyranny of discounting

Even for near-term events, discount rates can have a significant effect on our decisions. The example in Table 2.2, involving the levelized cost of electricity, shows that for an investment with a maximum lifetime of 45 years, the relation between the present value of two alternatives changes significantly when the annual discount rate changes from 2% to 4%. Discounting can be even more important when considering distant events.

The "tyranny" of discounting refers to the fact that, at non-negligible discount rates, events in the distant future have almost no effect on current decisions: the present discounted value of a cost, measured in either utility or dollars, in the distant future can be very small, even if the absolute cost in the future is very large. As a consequence, people today may not want to incur even a small current cost to avoid a large future cost. The logic of discounting "compels" us to essentially ignore the consequences of our actions on people in the distant future. Here we emphasize utility discounting, so the relevant discount rate is ρ, but the same logic applies to consumption discounting. A larger value of ρ implies that the planner is more impatient with regard to future utility: she is willing to give up less current utility in order to obtain an extra unit of future utility.

Environmental policy provides a kind of insurance. A person who buys standard insurance makes a fixed payment (the premium) that entitles her to a payout under certain contingencies. People can buy insurance against some natural events, such as floods or earthquakes, but society as whole cannot obtain insurance against a world-wide occurrence of climate change: there is no cosmic insurer standing outside our world, able to make a contract exchanging premiums for a payout in the event of a bad outcome. However, society can decide to incur near-term costs, e.g. by replacing fossil fuels with a more expensive alternative, to reduce the likelihood or the severity of future climate-related damages. These policies are analogous to insurance because they involve current costs (the "premium") to mitigate the consequences of unknown future events.

Climate-related damages associated with current emissions might not arise for many decades, or even centuries, but if they do occur they are likely to persist a long time. Our illustrative model incorporates both delay and persistence. Suppose that in the absence of costly changes (e.g. moving toward low-carbon energy) an "event", such as the melting of the Western Antarctic Ice Sheet (WAIS) will occur in 200 years, and will result in a loss of 100 units of utility in each subsequent period.[2] Society can avoid this event by paying, in perpetuity, a "premium" of z. The payment z is not literally an insurance premium; it is the flow cost of taking actions that eliminate the event, e.g. using expensive alternatives to carbon-based energy.

Figure 19.1 illustrates this scenario. The solid step function shows the trajectory of utility if society does not pay the premium: utility falls by 100 units, from 150 to 50, at the event time $t = 200$. If society can avoid the loss by paying a premium $z = 13.5$, the dashed line (constant at $150-13.5$) shows its utility trajectory. If, instead, society can eliminate this loss by paying a premium of only $z = 0.25$, the dotted line (constant at $150 - 0.25$) shows its utility trajectory. The largest premium society would be willing to pay, denoted Z, makes society indifferent between the trajectory shown by the solid step function, and the trajectory with constant utility $150 - Z$. We can compare these two trajectories by comparing the present discounted value of costs under them. If society pays the premium ,z, in every period, the present discounted value of the premium cost is $\int_0^\infty ze^{-\rho t}dt = \frac{z}{\rho}$. If society does not pay the premium, it incurs no cost until $t = 200$, and thereafter occurs the cost 100 in every period, leading to a present discounted cost of $\int_{200}^\infty 100e^{-\rho t}dt = e^{-200\rho}\frac{100}{\rho}$. The *maximum* premium that society would pay

[2] Such an "event" would actually occur over long periods of time, possibly centuries, not at a single point in time, as in our model.

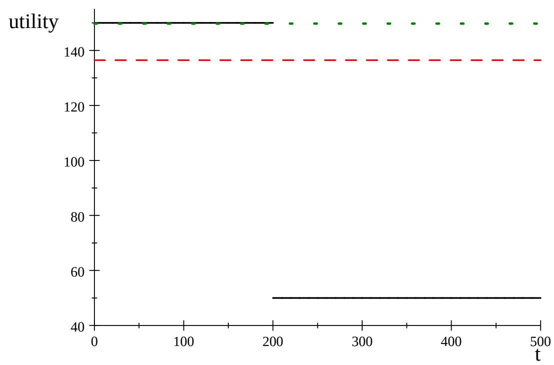

Figure 19.1: Solid step function: the utility trajectory when the event occurs at $t = 200$ and leads to a 100 unit drop in utility. Dashed and dotted lines show the constant utility trajectory when society pays a premium (13.5 or 0.25) that eliminates the event.

equates these two costs: $Z(\rho)$ is the solution to $e^{-200\rho}\frac{100}{\rho} = \frac{z}{\rho}$, so $Z(\rho) = e^{-200\rho}100$. Society is willing to pay any premium less than or equal to $Z(\rho)$ to avoid the loss beginning at $t = 200$.

The solid graph in Figure 19.2 shows this premium as a function of ρ, for an event time $T = 200$, illustrating the tyranny of discounting. At a discount rate of 1%, the maximum premium is about 13.5% of the loss, but at a discount rate of 3% the premium falls to less than 0.25%. With discounting at a non-negligible level, decisionmakers value the (finitely long) near future vastly more than they value the (infinitely long) distant future. Here, society has little incentive to incur even modest current costs, associated with climate change policy, to avoid large future costs.

19.1.2 Uncertain timing

The example above assumes that the time of the event is certain. Random timing likely increases the risk premium, because a deterministic model tends to understate actual costs arising under uncertainty. If, for example, there is a 50% chance that, in the absence of climate policy, the event will occur in 150 years, and a 50% chance that it will occur in 250 years, then the expected time of occurrence, 200 years, equals the certain time in the example above. The present value cost of the event is much greater if it occurs in 150 years,

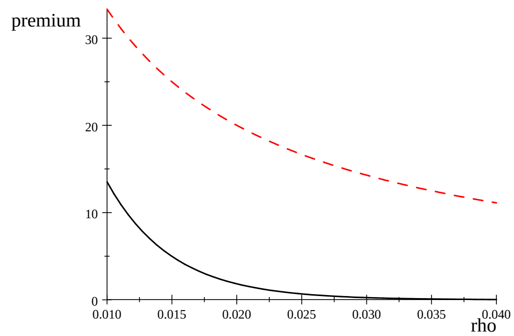

Figure 19.2: Maximum premium as a percent of the flow loss when *certain* event time is $T = 200$ (solid) and when event time is exponentially distributed with $E(T) = 200$ (dashed).

and only slightly less if it occurs in 250 years, both relative to the present value if it occurs in 200 years. Therefore, the *expected cost* (the probability-weighted average of the two costs) is closer to the higher, earlier cost: the expected costs in the stochastic scenario is greater than the known costs in the deterministic scenario.[3]

The dashed graph in Figure 19.2 shows the maximum risk premium if the time of the event is random (and exponentially distributed), with expected event time $T = 200$ (so that the two graphs are comparable).[4] Moving from the deterministic to the stochastic setting increases the maximum risk premium by a factor of 2.5 at $\rho = 0.01$, and by a factor of 57 at $\rho = 0.03$. Mistakenly treating the event time as deterministic, when in fact it is stochastic, can lead to a moderately large underestimate of the amount society should be willing to spend to avoid the event (the maximum risk premium) at small discount rates ($\rho = 0.01$), and a very large underestimate at higher discount rates ($\rho = 0.03$).

In the real world, stochasticity is important. We do not know if an event such as the melting of the WAIS will happen sooner or later, or perhaps

[3]This result is a special case of "Jensen's inequality": if T is a random variable, and $Z(T)$ is a convex function of T, then $E(Z(T)) > Z(E(T))$, where $E(\cdot)$ denotes "expectation". The present value, $\exp(-\rho T)$, is a convex function of T.

[4]If the probability that the event will occur over the next small unit of time, "dt", given that it has not yet occurred, is approximately $h \times dt$, with $h > 0$ a constant, then the event time is exponentially distributed; h is known as the "hazard rate", and the expected time of the event is $\frac{1}{h}$. For Figure 19.2, $h = \frac{1}{200} = 0.05$.

never (in a span relevant to human existence). If we ignore our uncertainty about the time of the event, and instead merely replace the random time by its expectation, we may vastly understate the amount that we should be willing to spend to prevent the event from happening. There are many other types of uncertainty. Our example assumes that the cost of the event is known, but both the cost and the timing are uncertain. The example also assumes that payment of the premium eliminates the possibility of the event. However, at best, costly actions such as the reduction of emissions decrease but do not eliminate the possibility of future climate events. The general point is that using a deterministic model (one that ignores uncertainty) to approximate a stochastic world can lead to large errors in formulating policy prescriptions. Stochastic models have only recently become widely used in climate economics.

Box 19.2 A different perspective on the tyranny of discounting. The present discounted value of a perpetual annual loss of x, equals the sum of the loss for the next 200 years ($\frac{1-e^{-\rho 200}}{\rho}x$) and the loss for the infinitely many years beginning 200 years from now ($\frac{e^{-\rho 200}}{\rho}x$). The ratio of these two losses, $\frac{e^{-\rho 200}}{1-e^{-\rho 200}}$, equals the value, to the decision-maker today, of the infinitely many years starting 200 years from now, relative to the value of the next 200 years. At a 1% discount rate this ratio is 0.157 and at a 3% discount rate the ratio is 0.0025. At 3% discounting, the planner values utility during the next 200 years (about 10 generations) 400 times as much as she values utility for the infinitely many years (and generations) beginning in 200 years.

19.2 The consumption discount rate

Objectives and skills

- Understand the relation between discounting utility and discounting consumption, and the Ramsey formula for the Consumption Discount Rate (CDR).
- Understand why growth has a large effect on the CDR, and thus on policy prescriptions.

For policy applications, the consumption discount factor, and the associated consumption discount rate, is more useful than the utility discount factor and rate. The SCC is computed using the consumption discount rate. Most people, including policymakers, care about consumption, income, jobs and the other things that produce utility, not utility itself. Asking a policymaker how many units of utility society should sacrifice today to obtain an extra unit of utility at some time in the future, will elicit a blank stare. (Readers know that the answer is "The utility discount factor".)

Asking the policymaker "*How many dollars of consumption should society be willing to give up today in order to obtain one extra dollar of consumption at a future time?*", is at least an intelligible question. The question captures the trade-off arising with policy that has costs and benefits at different points in time, such as climate policy. That policy may reduce consumption today, by requiring greater expenditures on pollution abatement or the switch to more expensive types of energy. By protecting the climate, the policy may make people in the future better off.

19.2.1 The Ramsey formula

The Ramsey formula shows the relation between the consumption discount rate, r, (CDR) and the utility discount rate, ρ (Appendix M.1)

$$\begin{gathered}\text{Ramsey formula for CDR: } r(t) = \rho + \eta_t g_t, \\ \text{using the definitions} \\ \eta_t \equiv -\frac{u''(c_t)}{u'(c_t)} c_t \text{ and } g_t \equiv \frac{\frac{dc}{dt}}{c_t} \text{ (the growth rate).}\end{gathered} \tag{19.1}$$

This formula shows that the CDR may change over time, but we first discuss the case where it is constant. To achieve this simplicity, we assume: (i) Utility equals $u(c) = \frac{c^{1-\eta}}{1-\eta}$, so that $-\frac{u''(c_t)}{u'(c_t)} c_t = \eta$, a constant. (ii) The growth rate for consumption is a constant, g. With these assumptions, the CDR is a constant, $r = \rho + \eta g$. The parameters have the following interpretation:

> If ρ is larger, the planner is less patient, and thus places less weight on future utility.
>
> If η is larger, the planner is more averse to inequality, and thus less willing to impose costs on one generation in order to benefit a richer generation.
>
> If g is higher, growth is faster, making people in the future that much richer than people today.

Anything that increases r, lowers the consumption discount factor, e^{-rt}, thus reducing the amount society should spend today to avoid a dollar loss in consumption at t. The parameter ρ is a measure of our impatience with respect to future utility. An increase in ρ makes future utility less valuable from the standpoint of today, thereby making a planner willing to forgo less current consumption (and utility) in order to obtain higher future consumption (and utility). Thus, an increase in ρ causes r to increase.

The parameter η is an inverse measure of society's willingness to transfer income from one point in time to another (the inverse of the elasticity of intertemporal substitution). A larger value of η (a smaller elasticity of intertemporal substitution) means that people are less willing to transfer income from one period to another. If we think of consumption at different points in time as corresponding to consumption for different people, then η provides a measure of aversion to inequality: a larger value of η means that society has a greater aversion to inequality.

Figure 19.3 helps to visualize the role of η in determining the willingness to move consumption from one period to a different period (or from one person to a different person). Each curve shows the combination of consumption levels in two periods (or for two people), denoted C and c, that lead to a constant sum of utility, $u(c)+u(C)$. The dashed curve corresponds to $\eta = 2$ and the solid curve corresponds to $\eta = 0.5$. The two curves are tangent, and represent the same sum of utility, at $c = C = 5$. [5] This figure abstracts from the role of impatience by setting $\rho = 0$.

Suppose that a utilitarian planner wants to maximize this sum of utility. The planner with $\eta = 2$ is indifferent between $(c, C) = (5, 5)$ (the point of tangency) and $(c, C) = (3.34, 10)$ (shown as point X in the figure); this planner is willing to give up 1.66 units of c in order to increase C by 5 units. The planner with $\eta = 0.5$ is indifferent between $(c, C) = (5, 5)$ and $(c, C) = (1.71, 10)$ (shown as point Y in the figure); this planner is willing to give up 3.39 units of c in order to increase C by 5 units. The larger is η, the more averse is the planner to inequality in consumption between the two periods (or two people). *The larger is η, the less consumption the*

[5] Each of the graphs is analogous to an isoquant; instead of showing combinations of factors of production leading to a constant level of output, the graph shows the combination of consumption levels, in the two periods, leading to a constant level of total utility. A normalization results in the two curves in Figure 19.3 being tangent at $C = c = 5$. This normalization is unimportant for our purpose here, which involves only the relative curvature of the two graphs, not their relative levels.

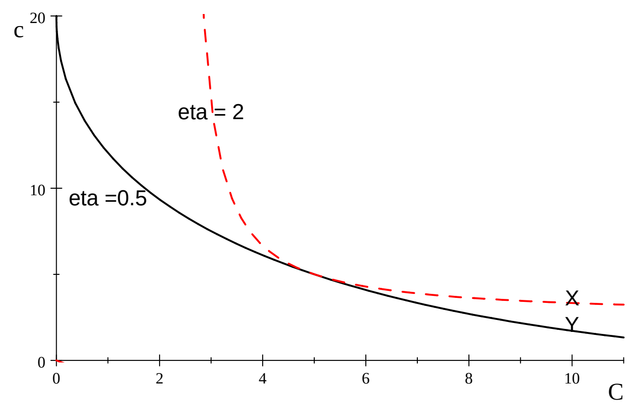

Figure 19.3: Combinations of consumption in two periods, C and c, that lead to a constant sum of utility, $U(c)+U(C)$, for $\eta = 0.5$ (solid) and $\eta = 2$ (dashed).

planner is willing to take away from a poorer person in order to increase the consumption of a richer person.

With decreasing marginal utility (a concave utility function) each additional unit of consumption provides a smaller increase in utility: a rich person values an extra \$100 less than a poor person does. A positive value of g (= growth) means that people in the future are getting richer, decreasing their marginal valuation of still higher consumption. Thus, a larger positive value of g (faster growth) makes the planner today less willing to sacrifice consumption today (when society is poorer) in order to increase consumption for the richer future. Larger growth increases the consumption discount rate (decreases the consumption discount factor). An increase in g has a greater effect on the consumption discount rate, the more averse the planner is to income inequality (the larger is η).

In summary, a planner has a higher consumption discount rate (lower consumption discount factor) the more impatient she is with regards to future utility (the larger is ρ), or the more rapidly people in the future are getting richer (the larger is g), or (for $g > 0$) the more averse she is to income inequality (the larger is η).

19.2.2 The importance of the growth trajectory

If ρ, g, and η are constant, then the CDR is also constant, and the insights from Chapter 19.1 apply, replacing "utility" with "consumption". The parameters ρ and η measure preference characteristics: the planner's impatience (ρ) and aversion to (intertemporal) inequality (η). Those parameters might change over time or with levels of consumption, but they are often treated as constants. The growth parameter, g, in contrast, describes the economy, not preferences; there is no reason to think that it is constant. Growth rates over long spans of human history have been close to zero, but growth rates over the past two centuries have been around 1.5% – 2%. Given the consensus view that η is not close to zero, a positive value of g has a significant effect on the CDR, and thus on society's willingness to sacrifice consumption today in order to protect future generations from climate damages.

If we expect high growth to continue over future centuries, then our successors will be much richer than we are; if we are somewhat averse to intergenerational income inequality (η is not close to 0), then it makes sense for us to be reluctant to incur costs in order to protect our much richer successors from (non-catastrophic) climate-related damages. This view relies on the assumption that growth over the next several centuries will resemble growth of the past two centuries, not growth over the previous millennia. It is not reasonable to assume that growth will abruptly stop, but it may be presumptuous to make long-lasting decisions based on optimism about growth. Growth experts were asked for their assessment of likely growth over the next two centuries. Most anticipate growth in the 1% – 3% range, but some expect negative growth and one expects growth above 6%. There appears to be little consensus amongst experts about future growth.

Appendix M.2 presents a model in which growth starts out at 2% and falls to zero gradually over time. We compare society's willingness to pay to avoid damages that begin in T years when it correctly anticipates this growth trajectory, versus when it is either "falsely pessimistic" (believing incorrectly that future growth will always be zero) or "falsely optimistic" (believing incorrectly that future growth will always be at 2%). False optimism makes society willing to spend too little, and false pessimism makes society willing to spend too little to avoid the future damages. Which of these errors is greater in magnitude? If the damages begin soon (T is small), then near-term growth is important. Our assumption that actual growth falls slowly means that the falsely optimistic view is closer to being correct in the near term, compared to the falsely pessimistic view. In this circumstance (T small), the error under pessimism is greater than under optimism. The reverse holds if damages begin in a century (T large), when the pessimistic view about growth is closer to being correct. For large T, the error under false optimism is much greater than under false pessimism.

The main effects of climate change are likely to occur a century or more in the future: T is large. The model suggests that the error we make in being too optimistic about growth (spending too little to avoid climate damages) is likely to be much greater than the error we make in being too pessimistic about growth (spending too much to avoid damages). This conclusion favors the use of caution (erring on the side of safety) in setting climate policy.

19.2.3 Growth uncertainty

Economists have followed two principal strategies to incorporate uncertainty, including uncertainty about growth. The first, most straightforward and widely used approach, changes the planner's welfare criterion to discounted *expected* utility (DEU). This criterion also adds up the discounted utility in different periods, but now takes expectations of the sum with respect to future consumption (or whatever variable is random). The second strategy for incorporating uncertainty replaces the model of DEU with a more general alterative.

Both approaches produce a "certainty equivalent" CDR, that can be used to evaluate how much society should be willing to invest today, in order to increase consumption (reduce damages) in an *uncertain* future world. In both cases, the certainty equivalent CDR generalizes the Ramsey formula 19.1; it involves the parameters of the distributions of the random variables, and parameters that describe attitudes to impatience, risk and intertemporal consumption transfers. The name "certainty equivalent" means that the discount rate can be used to assess a public investment as if the world were non-random; the randomness is already built in to the certainty equivalent discount rate.

Discounted expected utility The simplest modification replaces deterministic constant growth, g, with a random process, $\tilde{g}(t) = \bar{g} + \tilde{\varepsilon}_t$ where $\bar{g}$ equals expected growth and $\tilde{\varepsilon}_t$ is a serially uncorrelated mean-zero random variable. The absence of correlation means that growth in one period does not affect growth in subsequent periods. This model of uncertainty increases the amount we are willing to spend to avoid future damages, but not by much.

Positive correlation between growth in different periods increases this correction (lowers the CDR by more), because positive correlation makes the future riskier. An example, in which a person might get \$0 or \$1 in each of two periods, makes this relation intuitive. In one scenario, the probability of receiving either amount is the same, and the amount received in the first period has no effect on the likelihood of receiving a dollar in the second: the gifts are uncorrelated. In this case, the person receives a total of \$0 or \$2, each with 25% chance, and \$1 with 50% chance. In the other scenario, the person has an equal chance of receiving \$0 or \$1 (as in the first scenario) but the gifts are perfectly positively correlated: she receives the same amount in both periods. In this scenario, she has a 50% chance of obtaining either \$0 or \$2. In expectation, she obtains the same amount in both scenarios (\$1), but the variance of total receipts is higher in the second scenario; that scenario is riskier. Therefore, positively correlated random growth increases the amount society is willing to spend today to avoid future damages by substantially more, compared to uncorrelated random growth.

A richer model of stochastic growth uses

$$\tilde{g}(t+1) = \bar{g} + \alpha(\bar{g} - g(t)) + \tilde{\varepsilon}_t \tag{19.2}$$

where the shocks $\tilde{\varepsilon}_t$ are serially uncorrelated and $\bar{g}$ and α are parameters. The specialization $\bar{g} = 0$ and $\alpha = -1$ implies that growth is a random walk: growth in the next period equals current growth plus a random variable. The specialization $0 < \alpha < 1$ implies that growth is mean-reverting, approaching its long-run level $\bar{g}$; if the current growth is above $\bar{g}$, growth in the next period is expected to be less than current growth. Other models, involve longer lags or different assumptions about the distribution of the shock.

An alternative allows for the possibility of large ("catastrophic") shocks. For example, the shock $\tilde{\varepsilon}_t$ might be the sum of two random variables; the first has a "typical" (perhaps bell-shaped) distribution, and the second equals a large negative number with small probability and is otherwise zero. The realization of the second part of the random variable is zero in most years, but a catastrophe such as a world war sharply reduces growth (and income).

Using a model of stochastic growth to calculate the certainty equivalent CDR requires estimating the parameters of a model like equation 19.2. Researchers might then proceed as if the estimated model is the truth, i.e. as if the growth process really has the hypothesized form, with actual parameters equal to the estimated parameters. An alternative assumes that the function describing growth is known, but recognizes that the parameters are only estimated. This alternative, using Bayesian methods, models the evolution of future parameter estimates. The alternative provides a model both of how growth changes, and how our estimates of future growth change.

These alternatives (positively correlated growth, the possibility of catastrophes, Bayesian models that update parameter estimates) lead (almost always) to further reductions in the certainty equivalent CDR, increasing estimates of the amount that society should spend today in order to reduce damages in an uncertain future world. Implementing these modifications requires using data to estimate the models. These more sophisticated alternatives improve on naive models that assume deterministic growth, but they are still based on the premise that the distant future will look like the recent past; without that premise, growth data would be useless.

These methods of estimating the certainty equivalent CDR consider a single type of uncertainty, often uncertainty about growth. A different extension focuses on the correlation between growth and climate-related damages. Climate policy is an investment, requiring higher costs and reduced consumption due to the use of more expensive energy sources; the payoff of this investment is a reduction in future damages, and a corresponding increase in future consumption. The policy therefore indirectly creates a consumption transfer from today to future periods. Because growth is uncertain, we do not know the level of future consumption, absent this transfer. We therefore do not know the marginal value, to the future, of an additional unit of consumption. In addition, because of all of the uncertainties of climate science, we do not know how current policy would change the magnitude of future climate damages. Therefore, the "return on investment" of current climate policy is a random variable.

An investor deciding on how to allocate funds between a "market portfolio" (e.g. an index fund) and a particular stock, faces a similar problem. Box 6.1 sketches the idea that a stock that is negatively correlated with the market return provides a hedge against market risk, and therefore might be worth buying even if its expected return is below the expected market return. If climate policy is likely to yield a large return (reduce future damages by a large amount) in circumstances where the future is relatively poor, then climate policy provides a hedge against future growth uncertainty. In this case, investing in climate policy may be economically rational even if its expected return is below that of other social investments. In contrast, if climate policy is likely to provide a high return in circumstances where the future is rich, then climate policy should be required to pass a more stringent cost-benefit test, compared to other social investments. Currently, there is no consensus about which of these two possibilities is more likely.

A different paradigm The stochastic extensions of the Ramsey formula described above use discounted *expected* utility (DEU), adding up the discounted utility in different periods, and taking expectations with respect to the random variables. For decades economists have been aware that important implications of DEU are inconsistent with stock market data. The risk premium equals the difference between the expected return on a risky asset (e.g. a portfolio of stocks), and the return on a riskless asset such as US government bonds. For long periods, this risk premium has exceeded 6% in the US, and has also been high in other countries. Explaining this difference using DEU requires a value of η much larger than consensus estimates. This inconsistency is known as the equity premium puzzle.[6]

Attempts to resolve this puzzle within the framework of DEU use some of the extensions discussed above, including models of catastrophic events or learning about uncertain parameters. A different approach replaces the DEU model with "recursive utility". This alternative has enough free parameters to be made consistent with market data.

DEU uses a single parameter, η, to represent two characteristics of preferences. In the deterministic framework, η represents the inverse of the elasticity of intertemporal substitution ("inequality aversion"). In the stochastic framework, η also represents risk aversion. That is, η represents both the decision-maker's attitude to transferring consumption over different time periods (or different people), and also her attitude about transferring consumption over different "states of nature", corresponding to different realizations of a random variable. There is no reason why risk aversion and aversion to intertemporal transfers should be governed by the same parameter. The DEU model is too parsimonious, using one parameter to represent two logically different characteristics. Recursive utility is more general, permitting the distinction between these two preference characteristics.

[6]Resolving this puzzle by simply assuming that the actual value of η is much larger than consensus estimates, leads to the "risk-free rate puzzle": the conclusion that the riskless rate is much higher than observed rates.

A related objection to DEU can be explained using an example. Consider a trajectory consisting of only two periods, with no impatience regarding utility (a utility discount factor equal to 1). Consumption might be low or high, yielding low or high utility u^L or u^H respectively. In one deterministic scenario an agent obtains first high and then low utility, $\left\{u^H, u^L\right\}$, and in a second deterministic scenario she receives $\left\{u^L, u^H\right\}$. Because she is not impatient, and faces no uncertainty, the discounted utilitarian assigns the same payoff, $u^H + u^L$, to both scenarios; she is indifferent between them. Consider a third scenario in which the agent faces a lottery. With probability 0.5 she obtains $\left\{u^H, u^H\right\}$ and with probability 0.5 she obtains $\left\{u^L, u^L\right\}$. This agent faces intertemporal risk: she might have two good periods or two bad periods. Discounted expected utilitarianism evaluates this payoff by taking expectations over the random payoffs, assigning the value $0.5\left(u^H + u^H + u^L + u^L\right) = u^H + u^L$ to this lottery.

This example shows that the DEU model implies that the social planner, or the people she represents, are indifferent about intertemporal fluctuations in utility. Models of recursive utility include an additional parameter that measures intertemporal risk aversion. A planner who is intertemporally risk averse prefers the trajectory $\left\{u^H, u^L\right\}$ (equivalently, $\left\{u^L, u^H\right\}$) to the lottery over trajectories. With empirically plausible levels of intertemporal risk aversion, stochastic growth might have little effect on the consumption discount rate. That is, taking into account intertemporal risk aversion, and also recognizing that future growth is stochastic, can lead to a certainty equivalent consumption discount rate close to the (deterministic) consumption discount rate under zero growth. The policy implication is that stochastic (as distinct from zero) growth might lead to only small reductions in the amount an intertemporally risk averse planner is willing to spend today in order to avoid future damages.

19.3 Patience and intergenerational transfers

Objectives and skills

- Understand the meaning of and rationale for hyperbolic discounting.
- Understand the meaning and cause of time inconsistency.
- Understand the effect of hyperbolic discounting on climate change policy.

In order to discuss the role of patience and intergenerational transfers, we now abstract from both deterministic and stochastic growth, setting $g = 0$. However, we allow the patience parameter to depend on time, replacing ρ with $\rho(t)$.[7] The number of units of utility a person (or the planner who represents her) would give up at time 0 in order to obtain one extra unit of utility at time t depends on the *average* utility discount rate between time 0 and t; the utility discount factor is now $e^{-\int_0^t \rho(\tau)d\tau}$.

19.3.1 Explanation of hyperbolic discounting

We consider the case where $\rho(t)$ decreases with t, known as "hyperbolic discounting". (The case where $\rho(t)$ increases over time is empirically less interesting, but involves similar analysis.) We discuss hyperbolic discounting in four contexts: where a decision affects a single person or generation; where a decision creates transfers across different generations; where it creates transfer both within and across generations; and finally, a "physical" interpretation of hyperbolic discounting. We then explain its relevance to environmental and resource policy.

Transfers affecting a single person or generation Hyperbolic discounting provides a model of "excessive procrastination": deferring unpleasant tasks longer than we would like to. For example, suppose that we are told that a project due on December 10 will take five hours to accomplish if done on December 9, and only four hours if done on December 8. Because the project requires work (disutility), we prefer to put if off as long as possible; but we also prefer to spend as little time as possible on it, so there is a trade-off. On September 1 suppose that we can make a provisional plan to do the project on either December 8 or 9. It would be nice to delay for an extra day, increasing by 1% the amount of time we can put off the work. However, this extra 1% delay requires a 25% increase in the amount of time we will have to work when the day of reckoning arrives. The 25% extra work may seem more important (salient) than the 1% additional delay, leading us to decide, on September 1, to do the project on December 8.

---

[7] The argument t is the *distance* from the current calendar time, normalized to time 0, to a future time, t. It is not "calendar time". Thus, $t = 40$ regardless of whether, at calendar time 2020 we are considering an event at 2060, or whether, at calendar time 2050 we are considering an event at 2090.

Suppose that on December 8, a minute before we are scheduled to begin, we can reconsider our earlier plan. A postponement (doing the project on December 9 instead of December 8), gives us a 1441 minute delay, instead of the one minute delay if we carry out our original plan; it still requires a 25% increase in the amount of time needed to work. From the standpoint of December 8, the 144,100% increase in delay might be more salient than the 25% increase in working time, causing the person to reverse her earlier decision. Hyperbolic discounting can explain why a person changes the September 1 plan, and now delaying the project an additional day ("excessive procrastination").

A particularly simple form of ("quasi") hyperbolic discounting represents this situation using two "time preference" parameters, $0 < \beta \leq 1$ and $0 < \delta < 1$. The utility discount factor for $t \geq 1$ periods in the future is $\beta\delta^t$. A constant utility discount rate corresponds to $\beta = 1$ and hyperbolic discounting corresponds to $\beta < 1$. Suppose that the disutility of working 4 hours is $D(4)$ and the disutility of working 5 hours is $D(5)$. On September 1, the present value disutility of doing the project on December 8 is $\beta\delta^{100}D(4)$ and the present value disutility of doing the project on December 9 is $\beta\delta^{101}D(5)$. On September 1, the person prefers to do the project on December 8 if $\beta\delta^{100}D(4) < \beta\delta^{101}D(5)$, i.e. if $D(4) < \delta D(5)$. Once December 8 arrives, the choice is between doing the project on that day, having disutility $D(4)$, or procrastinating, and having present value disutility $\beta\delta D(5)$. On December 8, the person procrastinates if $D(4) > \beta\delta D(5)$.

Plans are "time-inconsistent" if a person wants to change an earlier plan, despite having received no additional information since the original plan was made. Putting the two previous inequalities together, we see that the plan made on September 1 is time inconsistent if

$$\beta\delta D(5) < D(4) < \delta D(5). \tag{19.3}$$

This inequality requires $\beta < 1$, i.e. hyperbolic discounting.

With time inconsistency, the modeler has to decide what type of outcome is "reasonable". Suppose that inequality 19.3 holds, so that time inconsistency arises in our example. If the person has a "commitment device" on September 1 that somehow binds them to completing the project on December 8, the optimal plan will be carried out. For example, the person may commit to getting married on December 9, making it prohibitively expensive to procrastinate when December 8 arrives. It may be costly to construct a commitment device. Absent such a device, there is nothing to keep the person from procrastinating, and the reasonable outcome is for the project to be completed December 9.

More generally, when time inconsistency arises, and commitment devices are impractical, the equilibrium outcome can be obtained by thinking of the decision problem as a game amongst a "sequence of selves". The "self" at time t takes an action (e.g. deciding whether to work on the project), taking into account how future "selves" will behave.

Transfers across generations We may, for example, feel appreciably closer to our children than to our unborn grandchildren, but make little or no distinction between the 10'th and the 11'th future generation. In that case, we would be willing take less from our children in order to enhance our grandchildren's welfare, than we would take from the 10'th future generation to enhance the welfare of the 11'th generation. Hyperbolic discounting formalizes this type of intergenerational perspective.

Time inconsistency arises here for the same reason as in the single agent example, but here the game involves a sequence of generations instead of a sequence of selves. Table 19.1 provides an example that helps to understand this situation. The example uses the parameters $\beta = 0.7$ and $\delta = 0.9$, so $\beta\delta = 0.63$. The current date is $t = 0$.

There are two investment opportunities, A and B. Both yield the same payoff, an increase of one unit of utility at $t = 2$, but they have different cost structures. Investment A requires no action and therefore no costs today ($t = 0$), but it requires an investment costing 0.64 units of utility to the generation alive the next period, at $t = 1$. Investment B requires costly actions both in the current and the next period. These actions create a utility loss to generation $t = 0$ of 0.17 and a utility loss to generation $t = 1$ of 0.625.

From the standpoint of the generation at $t = 0$, the present discounted benefit exceeds the present discounted cost for both investments, but this generation prefers investment A. With investment A, discounted benefit minus costs for generation $t = 0$ equals $-\beta\delta\,(0.64) + \beta\delta^2 1 = 0.163\,8$ and with investment B this discounted benefit minus costs equals $-0.17 - \beta\delta\,(0.625) + \beta\delta^2 1 = 0.003\,25$. Thus, generation $t = 0$ would like to compel its successor to make investment A. If it is not capable of this compulsion, generation $t = 0$ is willing to begin investment B, provided that it believes that its successor would complete the project.

From the standpoint of generation $t = 1$, investment A's present discounted value of benefits minus costs is $-0.64 + \beta\delta 1 = -0.01 < 0$. Investment A does not pass the cost-benefit test for generation $t = 1$. In contrast, if generation $t = 0$ had started investment B, the present discounted value of benefits minus (remaining) costs, from the perspective of the generation at $t = 1$ is $-0.625 + \beta\delta 1 = 0.005 > 0$. Knowing that the generation at $t = 1$ would complete the investment B, but not undertake investment A, the generation at $t = 0$ chooses investment B.

utility change	$t = 0$	$t = 1$	$t = 2$	PDV at $t = 1$	PDV at $t = 0$
investment A	**0**	$\mathbf{-0.64}$	**1**	-0.01	0.163 8
investment B	$\mathbf{-0.17}$	$\mathbf{-0.625}$	**1**	0.005	0.00325

Table 19.1 Bold entries show flow benefits in different periods. PDV = "present discounted value" of future stream from standpoint of $t = 1$ and $t = 0$. $\beta = 0.7, \delta = 0.9$.

In this example, the equilibrium is for generation $t = 0$ to begin investment B, and generation $t = 1$ to complete it. Investment B is absolutely more expensive than A ($0.17 + 0.625 = 0.795$ instead of 0.64 undiscounted units of utility), and has a much lower present discounted value for generation $t = 0$ (0.00325 instead of 0.1638). However, in the absence of a commitment device, investment A is not feasible, whereas investment B is. This example illustrates the general point that with hyperbolic discounting, and lacking a credible device for committing future generations to act in a certain way, an earlier generation may chose a "less efficient" investment.

This example has parallels with climate change policy. Protecting the future from climate damage requires investment in low-carbon alternative energy supplies. From the standpoint of the current generation, the best policy may be to do nothing, requiring the next generation to undertake the entire cost of creating the low-carbon alternative (option A above). But the next generation possibly has the same incentive to delay. If the current generation delays, nothing is done in either period. However, by undertaking an expensive down payment on the low-carbon technologies (option B), the current generation may be able to change the trade-off that the next generation faces, inducing that generation to complete the investment needed to protect against climate damage.

Transfers within and across generations Climate policy likely creates transfers both within and across generations. For example, a switch to low-carbon fuels that decreases the (current) utility of those currently alive may benefit some of these people late in their life, and may also benefit people who have not yet been born. The first is an intra-generational transfer (from a person at one point to another point in their life) and the second is an inter-generational transfer. People might discount their *own* future utility at a constant rate and also discount the utility of future generations at a constant rate. However, there is no reason to think that they would use the *same* constant rate to discount their own and future generations' utility. Agents might discount these distinct types of transfers at different rates.

The distinction between intra- and inter-generational transfers requires an "overlapping generations model", one that recognizes that at any point in time some people are old and some are young; over time, the old die, the young become old, and new youngsters are born. If people in an overlapping generations framework use a lower discount rate to evaluate inter-generational transfers compared to intra-generational transfers, their preferences exhibit hyperbolic discounting.[8] For example, a person might be impatient for their own future utility, and also be a "luck egalitarian", unwilling to disadvantage future generations merely because of the date of their birth.

A "physical" interpretation of hyperbolic discounting Ramsey (1928) remarked "My picture of the world is drawn in perspective. ...I apply my perspective not merely to space but also to time." Perspective applied to both space or time seems to be part of our cultural DNA. Perspective implies that objects further in the distance, either spatially or temporally, appear smaller. Perspective applied to time implies that a unit of utility in the future appears less valuable than a unit of utility today: the utility discount factor decreases with time. A declining discount factor requires that the utility discount rate, ρ, is positive, but tells us nothing about whether ρ is constant.

The simplest model of *spatial* perspective, known as "one point perspective", can be visualized as railroad tracks that are parallel but which appear

[8]In the simplest overlapping generations model, the population is constant and people live for two periods. This model gives rise to the β, δ quasi-hyperbolic discounting discussed above.

to converge to a point, as they vanish into the distance. The actual railroad ties that connect these tracks each have the same length (because the tracks are parallel), but the ties appear to get smaller as they become more distant. The ratio of the *apparent* length of two successive ties (with the closer tie in the denominator), provides a spatial analog of the time discount factor. This ratio is always less than 1 (because the ties appear to be getting smaller) but it increases with distance. Therefore, the spatial analog of the discount rate falls with distance. To the extent that one accepts one point perspective as a model of spatial perspective, and also agrees that spatial and temporal perspective are analogous, hyperbolic discounting appears to be part of our cultural DNA.

19.3.2 The policy-relevance of hyperbolic discounting

Economists disagree about the policy-relevance of hyperbolic discounting. One basis for skepticism is that for intergenerational problems such as climate change, no utility discounting is ethical. If we require $\rho = 0$, there is no reason to be interested in the possibility that ρ decreases over time.[9] Our discussion of transfers both within and across generations illustrates the problem with this objection. One might agree that intergenerational transfers be discounted at rate 0, but recognize that people appear to be impatient for, and therefore discount, their own future utility. Why should social policy not take this intra-personal impatience into account? Allowing the inter-generational discount rate to be lower than the intra-generational rate results in hyperbolic discounting.

Rejection of hyperbolic discounting also puts the modeler on the horns of a dilemma. Setting ρ at a non-negligible positive constant leads to the tyranny of discounting discussed above. Setting ρ to a constant close to zero overcomes the tyranny of discounting, but it also implies (in DEU models) that current generations should be willing to save man-made capital at rates far in excess of those actually observed. Investment in man-made capital, particularly if it depreciates quickly, depends primarily on near-term discounting, whereas investment in long-lived natural capital, such as the climate, depends on discounting over long spans of time. Hyperbolic discounting offers an escape from this dilemma: models with a declining $\rho(t)$ can

[9] Instead of insisting that $\rho = 0$, a luck egalitarian might set it at a very small value to account for the possiblity that our species will be suddenly made extinct, e.g. by a comet striking the earth.

produce equilibrium savings rates equal to observed levels, while still leading to appreciable investment in climate policy. A small body of research uses hyperbolic discounting to model climate policy.

19.4 Summary

Climate policy (like natural resource policy in general) is a type of public investment, incurring costs with the expectation of future benefits. Economists base climate policy models on consensus views from climate science. Recognizing that resource policy involves trade-offs, most of these models involve discounting. Discounting renders costs and benefits in different periods and for different projects commensurable, and makes it possible to evaluate policies and to recommend the optimal level of carbon tax, or of other policies that reduce emissions.

Because we do not know the actual socially optimal level of climate policy, it is not possible to determine whether the carbon taxes recommended by mainstream economic models are too high, too low, or about right. To some environmental activists, a modest carbon tax seems inconsistent with the severity of the problem of climate change. Discounting is an important component of economic climate models, affecting the level of policy prescriptions. A higher discount rate corresponds to a lower discount factor, placing less weight on the future, and leading to more modest policy recommendations. Consumption discount rates (CDRs) depend on levels of impatience and "optimism".

If the economy grows during the next several centuries at rates seen during the last two centuries, people in the future will be much richer than we are, and able to tolerate the reduced consumption caused by non-catastrophic climate change. If we accept that the poor should not sacrifice in order to benefit the rich, and if indeed future generations will be richer than current generations, and furthermore we are confident that climate damage will be non-catastrophic, then society today should make no more than modest investments to protect the climate.

However, we do not know if growth will continue to be high during the next centuries, or if climate change will be non-catastrophic. The effect of growth uncertainty on the "certainty equivalent" CDR is sensitive to model specification. If we think of future growth as a sequence of serially uncorrelated random shocks (so that growth in one period does not affect growth in another), recognition of uncertainty leads to a small decrease in the certainty equivalent CDR, and a correspondingly small increase in the optimal level of environmental policy. Positive serial correlation of future growth increases the variance of future income. Variable income is "less valuable" than certain income, so positively correlated growth leads to a larger reduction in the CDR, and a larger increase in recommended policy.

The recognition that both economic growth and climate change are uncertain, further complicates matter. If a lower carbon stock is particularly valuable to future generations when growth has been relatively low, then growth and the return on the investment in climate policy are negatively correlated. In that case, climate policy provides a hedge against stochastic growth, making it easier to justify strict climate policy. If, however, a lower carbon stock is particularly valuable to rich future generations, the case for climate policy is weaker. The current stage of research has not reached a consensus on the sign of the correlation.

The standard paradigm uses a single parameter to measure a person's attitude to random income and to changes in income over time. A generalization, known as recursive utility, disentangles these two characteristics, leading to a different calibration and in some cases to significant differences in policy prescriptions. This generalization also incorporates a particular type of "intertemporal" risk aversion, which the standard paradigm of discounted expected utility assumes is zero.

The standard paradigm also treats the parameter that measures impatience with respect to future utility as a constant. A declining rate of impatience can explain "excessive" procrastination. More important for environmental policy, a declining rate also can distinguish between intra- and inter-generational transfers. Today's climate policy can effect the utility of currently living people late in their life, and also the utility of people not yet born. This climate policy therefore involves both intra- and inter-generational transfers. A declining rate of impatience often creates time-inconsistency, and requires solving a game instead of an optimization problem in order to assess environmental policy. A declining rate of impatience is consistent with high discounting in the short run, and low discounting in the long run. It can therefore reconcile observed savings rates (which depend primarily on near term discounting) with strong protection for the climate (which depends primarily on long term discounting).

19.5 Terms, study questions, and exercises

Terms and concepts

Positive and normative statements; Social Cost of Carbon, Discounted expected utilitarianism, utility discount rate and factor, consumption discount rate and factor, tyranny of discounting, hazard rate, Ramsey formula, elasticity of intertemporal substitution, inequality aversion, consumption growth rate, certainty equivalent discount rate, recursive utility, intertemporal risk aversion, equity premium puzzle, (quasi) hyperbolic discounting, intra- and inter-generational transfers, time inconsistency, overlapping generations, game among "sequence of selves".

Study questions

1. Explain the difference between discounting utility and discounting consumption.

2. Explain what it means to say that discounting is "tyrannical".

3. Explain the sense in which climate policy provides a kind of insurance.

4. Given the Ramsey formula for the consumption discount rate, explain the meaning of each term.

5. Explain why near-term growth is particularly important in evaluating a public investment project that has a payoff in the near term, whereas growth over long periods of time are important in evaluating a public investment project that has a payoff in the distant future.

6. Sketch some of the ways in which the certainty equivalent consumption discount rate the stochastic growth.

Exercises

Background for Question 4 Economists often work with the natural log of variables instead of their levels. This choice makes it easy to calculate growth rates and also means that in some applications we are able to work with the sum of two random variables, instead of their product, facilitating some calculations. Let C_2 and C_1 be the levels of consumption in two periods. The growth rate of consumption across these two periods is $G_2 = \frac{C_2 - C_1}{C_1}$, or

$G_2 + 1 = \frac{C_2}{C_1}$. Define $g_i = \ln(G_i + 1)$. To a first order approximation, $g_i \approx G_i$. Consumption in period 2 is $C_2 = (1 + G_2) C_1 = (1 + G_2)(1 + G_1) C_0$. Taking logs gives $\ln C_2 = \ln(1 + G_2) + \ln(1 + G_1) + \ln C_0$. Normalize by setting $C_0 = 1$, define $c_2 = \ln C_2$ and use the definition of g to obtain $\tilde{c}_2 = \tilde{g}_1 + \tilde{g}_2$. The tildas emphasize that these are random variables. The log of second period consumption equals the *sum* of two random variables.

1. Provide an intuitive explanation for the claim in the text that $WTP(\gamma, T)$ decreases in both γ and T.

2. (*) Write the formula for $WTP(\gamma, T)$ and confirm algebraically that it decreases in both γ and T.

3. For the example at the beginning of Chapter 19.2.3 (where income in two periods is either 0 or 1) calculate the variance of the payoff in the two cases, first where the two income levels are uncorrelated, and second where they are perfectly correlated.

4. Current consumption is $C_0 = 1$. Find the formulae for the expectation and the variance of a sum of two random variables (look it up in a statistics textbook or Google it). Use this formula and the "background" above, with $\ln c_2 = \tilde{g}_1 + \tilde{g}_2$. Given information at period 0, they have the same mean, $E\tilde{g}_1 = E\tilde{g}_2 = \bar{g}$, and the same variance, $var(\tilde{g}_1) = var(\tilde{g}_2) = \sigma^2$, and their correlation coefficient is ϕ. (The usual symbol for correlation is usuallyρ, but we have used that symbol for the utility discount rate.) Show how the mean and variance of $\ln C_2$ depends on ϕ.

Sources

Llavador et al. (2015) provide a detailed criticism of the discounted utilitarianism model, particularly as applied to climate change policy, and they suggest a sustainability-based alternative.

The examples in Chapter 19.1 are taken from Karp (2016).

Arrow et al (2012) examine discounting in an intergenerational context.

Gollier (2014) and Lemoine (2015) offer differing perspectives on the correlation between the market return and the return to protecting the climate

Gillingham et al. (2015) use a suite of models to examine uncertainty about climate change. They provide a survey of growth experts' assessment of future growth.

Gordon (2015) provides a comprehensive history of American growth, and explains why he expects that high growth rates during the last century will not continue into the future.

Mehra (2003) discusses the equity premium puzzle and reviews some of the explanations offered for it.

Epstein and Zinn (1991) provide an early empirical application of recursive utility.

Traeger (2014) explains the theory of intertemporal risk aversion and shows how it alters the relation between stochastic growth and the consumption discount rate.

Barro (2006) models the effect of rare catastrophic events on discount rates.

Laibson (1997) discusses the role of quasi-hyperbolic discounting in a model of savings.

Karp (2005) applies hyperbolic discounting to climate policy in partial equilibrium settings.

Ekeland and Lazrak (2010) note the relation between overlapping generations and hyperbolic discounting.

Ramsey (1928) is the source of the quote on time perspective.

Gerlagh and Liski (2012) and Iverson (2015) apply hyperbolic discounting to climate policy in a general equilibrium setting.

Appendix A

Ehrlich versus Simon

The Ehrlich-Simon bet involved prices of metals for which there are well established property rights. These metal prices have been volatile over the past century, and in one respect Ehrlich was simply unlucky. Had he picked a different decade in the 20th century, he would have had a better than even chance of winning the bet. The price of this basket of metals fell dramatically during the economic upheaval following World War I, and in most subsequent decades rose gradually. The volatility of metal prices has been largely due to macro economic cycles (recessions or booms) or political events (wars or boycotts) unrelated to scarcity. The theory developed in subsequent chapters explains why, putting aside these reasons for price volatility, modest but not spectacular price increases might be expected. The basis for this theory is that resource owners can "arbitrage over time", advancing or postponing their sales in order to take advantage of expected price changes. This arbitrage tends to lead to modest expected price increases.

Basing the wager on these metal prices was, in some respects, an odd choice for both the resource optimist (Simon) and pessimist (Ehrlich). It was an odd choice for Simon, the economist, because economic theory predicts modest price increases, not decreases. It was an odd choice for Ehrlich, the ecologist, because it involves the category of resources for which market forces are most likely to "work well": those having strong property rights.

Five years later, Ehrlich proposed a different bet, involving changes CO_2 concentrations, temperature, tropical forest area, and rice and wheat stocks per person, rather than commodity prices. Simon rejected Ehrlich's offer, and countered with a wager involving direct measures of human well-being, including life expectancy, leisure time, and purchasing power. Ehrlich declined that offer. Some of the components of these differing proposals reflect different views about the relation between mankind and natural resources. The resource pessimist begins with the premise that human welfare is inextricably linked to resource stocks; their degradation must eventually lower human welfare. For example, society may be well off during the period it uses fish or forest resources intensively, degrading their stocks. If resource use is unsustainable, resource extraction must eventually fall, and with it, human well-being. This overuse is more likely for resources where property rights are weak. The resource optimist, in contrast, starts from the premise that society will be able to find new resource stocks (new sources of fossil fuels) or alternatives to those resources (solar power instead of fossil fuels)

Appendix B

Math Review

This appendix reviews some concepts and results from basic calculus. It can be used as a reference during the course, and also gives readers an idea of the level of mathematics required for the course. The text assumes that readers have seen much this material before.

1. Derivatives and graphs The derivative of a function, $f(x)$ at a point x_0, written $\frac{df(x_0)}{dx}$, is the tangent ("slope") of a function, evaluated at a particular point, here x_0. If $f(x) = a + bx$, where a, b are independent of x (and therefore constants for our purposes here), then $\frac{df}{dx} = b$, a constant.

More generally, however, the value of the derivative depends on x. Figure B.1 shows the graph of $f(x) = 2 + 3x - 4x^2 - 5x^3$, the solid curve, the graph of $g(x) = \frac{df(x)}{dx}$, the dashed curve, and the graph of $h(x) = \frac{dg(x)}{dx} = \frac{d^2 f(x)}{dx^2}$. This figure reminds the reader that, in general, (1) a derivative is a function, not a constant; (2) where a function reaches an extreme point (a maximum or a minimum) the derivative of that function equals 0. If the graph of a function has an inflexion point, i.e. switches from being concave to convex, the second derivative of the function equals 0 at the inflexion point.

Another way to indicate that a function is being evaluated at a particular point, say $x = 2$, uses subscripts. For example, the subscript "$|x = 2$" here indicates that we evaluate the derivative of f with respect to x at $x = 2$:

$$\frac{df(x)}{dx}_{|x=2} = 2 + 3(2) - 4(2)^2 - 5(2)^3.$$

It is worth repeating that (in general) $\frac{df(x)}{dx}$ is itself a function of x; above we called this function $g(x)$. Similarly, $\frac{d^2 f(x)}{dx^2}$ is a function of x; above we

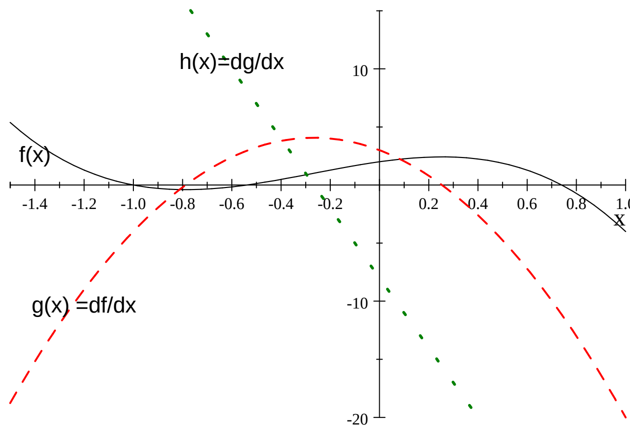

Figure B.1: The solid curve shows the graph of $f(x)$, the dashed curve shows the graph of $g(x) = \frac{df}{dx}$, and the dotted curve shows the graphof $h\left(x\right) = \frac{dg}{dx} = \frac{d^2 f}{dx^2}$.

call it $h\left(x\right)$.

Another way to write the derivative uses the "prime sign", $'$:

$$\frac{df\left(x\right)}{dx} = f'\left(x\right).$$

2. Derivatives of exponents. In the example above we took the derivative of a function involving an exponent. Students should know the following rule: if a is a constant (with respect to x), then

$$\frac{d\left(x^a\right)}{dx} = ax^{a-1}.$$

We write that "a is a constant with respect to x", instead of merely writing "a is a constant" because the formula above is correct even if a is a function of other variables (not x). For the purpose of taking this derivative, it does not matter whether a is a literally a constant or merely a constant with "respect to x". What matters is that a change in x does not change a.

3. The sum, product, and quotient rules. Students should know a few of the primary rules for derivatives. Suppose we have two functions of x, $a\left(x\right)$ and $b\left(x\right)$. (Note: in the previous line we treated a as a constant. Here we treat it as a function. In general, we are careful *not* to use the same symbol to mean two different things. Here, we intentionally use the same symbol, a, to mean two different things, first a constant and then a function. We want to encourage readers to pay attention to definitions.) We can form other functions using these two functions.

If c is the *sum of these two functions*, then

$$c(x) = a(x) + b(x) \text{ and } \frac{dc}{dx} = \frac{da}{dx} + \frac{db}{dx}.$$

The derivative of a sum equals the sum of a derivative. For brevity we write, for example, $\frac{da}{dx}$ instead of $\frac{da(x)}{dx}$.

If c is the *product of the two functions*, then

$$c(x) = a(x) \times b(x) \text{ and } \frac{dc}{dx} = \frac{da}{dx}b + \frac{db}{dx}a.$$

If c is the *quotient of two functions*, then

$$c(x) = \frac{a(x)}{b(x)} \text{ and } \frac{dc}{dx} = \frac{b\frac{da}{dx} - a\frac{db}{dx}}{b^2}.$$

4. The chain rule. The chain rule enables us to take the derivative of a function of a function. Suppose that y is a function of x and x is a function of z. Then y is a function of z, via the effect of z on x. The chain rule states

$$\frac{dy}{dz} = \frac{dy(x(z))}{dx}\frac{dx(z)}{dz}.$$

For example, if $y = x^{0.3}$ and $x = 7z$, then $\frac{dy(x)}{dx} = 0.3x^{0.3-1}$ and $\frac{dx}{dz} = 7$, so

$$\frac{dy}{dz} = \frac{dy(x(z))}{dx}\frac{dx(z)}{dz} = 0.3x^{0.3-1} \times 7 = 0.3(7z)^{-0.7} \times 7.$$

5.Partial derivatives. Some of our functions involve two arguments, instead of one. Throughout the book we use a cost function that depends on the stock of the resource, x, and the amount that is extracted in a period, y. We write this cost function as $c(x, y)$. A partial derivative tells us how the value of the function (here, costs) changes if we change just one of the variables, either x or y. We use the symbol ∂ instead of d to indicate that we are interested in the partial derivative.

We frequently illustrate concepts using the following specific cost function

$$\text{Parametric example: } c(x, y) = C(\sigma + x)^{-\alpha} y^{1+\beta},$$

where C, α, σ, and β are non-negative parameters. (Note that lower case c is a function, and upper case C is a parameter; this use of two similar symbols to mean two different things is intentional. It is important to pay attention to definitions.) The partial derivatives of this function with respect to x and y are

$$\frac{\partial C(x+\sigma)^{-\alpha}y^{1+\beta}}{\partial x} = -\alpha C\left(x+\sigma\right)^{-\alpha-1}y^{1+\beta},$$
$$\frac{\partial C(x+\sigma)^{-\alpha}y^{1+\beta}}{\partial y} = \left(1+\beta\right)C\left(x+\sigma\right)^{-\alpha}y^{\beta}.$$

Because we use this formulation throughout the text, the reader should be sure to understand it at this point. For example, in taking the partial of c with respect to x, we recognize that $Cy^{1+\beta}$ does not depend on x. Thus, in writing the partial of c with respect to x we treat $Cy^{1+\beta}$ as a constant. Although not literally a constant, this term is *constant with respect to x*. In English: this term does not depend on x; therefore, changes in x do not affect this term. To drive this point home, we can write

$$c\left(x,y\right) = \text{“Constant with respect to } x\text{”} \times \left(\sigma+x\right)^{-\alpha}$$

and then use the rule in item #2 above to write the partial derivative of c with respect to x as

$$\frac{\partial c(x,y)}{\partial x} = \text{“Constant with respect to } x\text{”} \times \frac{d(\sigma+x)^{-\alpha}}{dx}$$
$$= \text{“Constant with respect to } x\text{”} \times \left(-\right)\alpha\left(\sigma+x\right)^{-\alpha-1}$$
$$= -\alpha C\left(\sigma+x\right)^{-\alpha-1}y^{1+\beta}$$

The two partial derivatives of c with respect to x and y are themselves functions of x and y. Thus, we can differentiate either of these functions, with respect to either x or y, to obtain a higher order partial derivative. For example,

$$\frac{\partial^2 C(x+\sigma)^{-\alpha}y^{1+\beta}}{\partial y^2} = \frac{\partial\left[\frac{\partial C(x+\sigma)^{-\alpha}y^{1+\beta}}{\partial y}\right]}{\partial y} = \frac{\partial\left[(1+\beta)C(x+\sigma)^{-\alpha}y^{\beta}\right]}{\partial y}$$
$$= \left(1+\beta\right)\beta C\left(x+\sigma\right)^{-\alpha}y^{\beta-1},$$

and

$$\frac{\partial^2 C(x+\sigma)^{-\alpha}y^{1+\beta}}{\partial y \partial x} = \frac{\partial\left[\frac{\partial C(x+\sigma)^{-\alpha}y^{1+\beta}}{\partial y}\right]}{\partial x} = \frac{\partial\left[(1+\beta)C(x+\sigma)^{-\alpha}y^{\beta}\right]}{\partial x} =$$
$$\left(1+\beta\right)\left(-\alpha\right)C\left(x+\sigma\right)^{-\alpha-1}y^{\beta}.$$

6. Total derivatives. A function may depend on two variables, and each of those variables might depend on a third variable. The *total derivative* tells us how much the function changes for a change in this third argument. For example, suppose that costs depend on x and y, as above, and x and y both depend on ε. We show this dependence by writing $x(\varepsilon)$ and $y(\varepsilon)$. With this notation, we write costs as $c(x(\varepsilon), y(\varepsilon))$. The total derivative of c with respect to ε is

$$\frac{dc(x(\varepsilon), y(\varepsilon))}{d\varepsilon} = \frac{\partial c}{\partial x}\frac{dx}{d\varepsilon} + \frac{\partial c}{\partial y}\frac{dy}{d\varepsilon}. \tag{B.1}$$

In writing this equation, we merely apply the chain rule twice: the first term accounts for the fact that a change in ε alters c via the change in x, and the second term accounts for the fact that a change in ε alters c via the change in y. The total change in c due a change in ε is the sum of these two terms.

This expression might seem complicated, but most of the applications in this book are extremely simple. We will be interested in the case where $x = x_1 - \varepsilon$ and $y = y_1 - \varepsilon$, where x_1 and y_1 are treated as constants for the purpose here. For these two functions, we have

$$\frac{dx}{d\varepsilon} = -1 \text{ and } \frac{dy}{d\varepsilon} = -1. \tag{B.2}$$

Substituting equation B.2 into equation B.1 gives the total derivative

$$\frac{dc(x(\varepsilon), y(\varepsilon))}{d\varepsilon} = -\frac{\partial c(x_1 - \varepsilon, y_1 - \varepsilon)}{\partial x} - \frac{\partial c(x_1 - \varepsilon, y_1 - \varepsilon)}{\partial y}.$$

We often want to evaluate this derivative at $\varepsilon = 0$. In this case, we write

$$\frac{dc(x(\varepsilon), y(\varepsilon))}{d\varepsilon}_{|\varepsilon=0} = -\frac{\partial c(x_1, y_1)}{\partial x} - \frac{\partial c(x_1, y_1)}{\partial y}.$$

Remember that the subscript "$|\varepsilon = 0$" on the left side of this equation means that we evaluate the derivative of c with respect to ε where $\varepsilon = 0$.

7. Constrained optimization. Suppose that the problem is to maximize $V(z)$ ("value") subject to $0 \leq z \leq 4$. We might have either an interior equilibrium (a solution where $0 < z < 4$) or a boundary equilibrium (where $z = 0$ or $z = 4$). The function V represented by graph B has an *interior* optimum (the optimal z is between the two boundaries, 0 and 4). At this interior optimum, $\frac{dV}{dz} = 0$: a marginal increase or decrease in z does not change $V(z)$. The function V represented by both graphs A and C have *boundary* optima: $z = 0$ for A, where $\frac{dV}{dz} < 0$ and $z = 4$ for C, where $\frac{dV}{dz} > 0$.

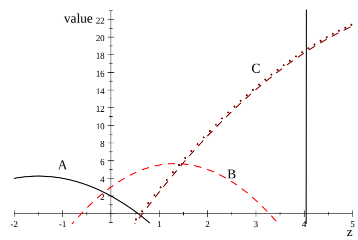

Figure B.2: The curve B has an interior maximum. The curves A and C have boundary maxima.

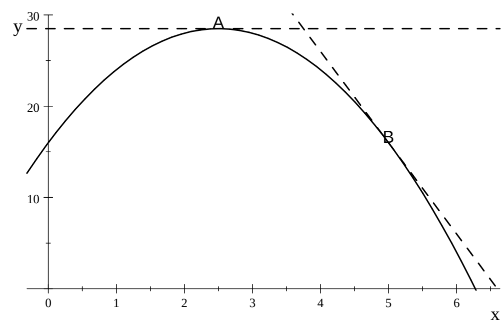

Figure B.3: A change in x away from point A has only a second order effect on y. A change in x away from point B has a first order effect on y.

For A a decrease in z at $z = 0$ increases V and for C an increase in z at $z = 4$ increases V, but either of these changes violates a constraint and thus is not feasible.

8. First and second order effects. Figure B.3 illustrates the meaning of "first order" and "second order" effects. The figure shows a graph, the solid curve, and the dashed tangencies at two points, the maximum point A and an arbitrary point B. The first derivative of the function at $x = 2.5$ (the horizontal coordinate of A) is zero and the second derivative is nonzero (negative). A very small movement away from $x = 2.5$ results in *negligible* change in the value of y: the first order effect, on y, of the change in x is zero, and the second order effect is negative. In contrast, a very small movement away from $x = 5$ (the horizontal coordinate of point B) results in a *non-negligible* ("first order") change in y, because the derivative of the function at $x = 5$ is nonzero.

Exercises

1. For the function $f(x) = -2x + x^{0.4}$, graph $f(x)$, $f'(x)$, and $f''(x)$. (A sketch is adequate – it does not have to be precise.)

2. Find the condition (an equation) for an extreme point of f. Is this extreme point a maximum or a minimum. (How do you know?)

3. For the function $g(x, y) = x^2 y^{0.2} - 3y$, (a) write the partial derivatives of g wrt x and y and (b) evaluate these derivatives at $x = 3$, $y = 1$.

4. Suppose that you are told $x = t^2$ and $y = 7t$. (a) Write the total derivative, wrt t, of $h(t) = g(x(t), y(t))$, where $g(x, y) = x^2 y^{0.2} - 3y$. Evaluate this derivative at $t = 1$.

5. Evaluate
$$\frac{d\left(\frac{4z^2+3z}{7z^{0.5}}\right)}{dz}_{|z=1}$$

6. Evaluate
$$\frac{d\left(\left(4z^2+3z\right)\left(7z^{0.5}\right)\right)}{dz}_{|z=1}$$

7. Suppose that demand is $D = 10 - 2P$, where P is price. (a) Evaluate the elasticity of demand at $P = 2$. (b) Evaluate the elasticity of demand at $P = 3$. (c) Write the marginal revenue, evaluated at these two prices.

8. Suppose that demand is $D = 10P^{-1.2}$. (a) Evaluate the elasticity of demand at $P = 2$. (b) Evaluate the elasticity of demand at $P = 3$. (c) Write marginal revenue, evaluated at these two prices.

9. For the constrained maximization problem
$$\max_{x,y}\ 4xy - 3x^2 - y^2$$
$$\text{subject to}\ \ x + 4y = 17,$$
Use the constraint to solve for x as a function of y. Substituting this result into the maximand (the object you are maximizing) write the first order condition for the optimal y.

Appendix C

Comparative statics

Comparative statics exercises often use differentials of an "equilibrium condition". This equation might be the first order condition to an optimization problem, or it might be a statement that says "supply equals demand". The equilibrium condition determines an "endogenous variable" (e.g., the optimal level of sales or the price that equates supply and demand) as a function of model parameters. A comparative statics exercise asks how the endogenous variable changes as one or more parameters of the model change.

The comparative statics exercises discussed here use the *differential* of a function. A function might depend on several arguments. Suppose that a function L depends on x, y, z: $L = L(x, y, z)$. The differential of L, denoted dL is

$$dL = \frac{\partial L}{\partial x}dx + \frac{\partial L}{\partial y}dy + \frac{\partial L}{\partial z}dz.$$

The change in L (denoted dL) equals the change in L due to the change in x, $\frac{\partial L}{\partial x}$ times the change in x, dx, plus the change in L due to the change in y, $\frac{\partial L}{\partial y}$, times the change in in y, dy, plus the change in L due to the change in z, $\frac{\partial L}{\partial z}$, times the change in z, dz.

To illustrate the use of differentials in comparative statics experiments, suppose that the demand is $Q^D = p^{-0.6}$ and supply is $Q^S = 2 + \beta p + p^{0.5}$, with $\beta > 0$. Define excess demand, $E(p, \beta)$, as demand minus supply. The equilibrium price equates supply and demand, i.e. it sets *excess demand* equal to 0:

$$E(p, \beta) = p^{-0.6} - \left(2 + \beta p + p^{0.5}\right) = 0.$$

For this example, we have one endogenous variable, p, and one parameter, β. We cannot solve the equilibrium price as a function of β. Figure C.1

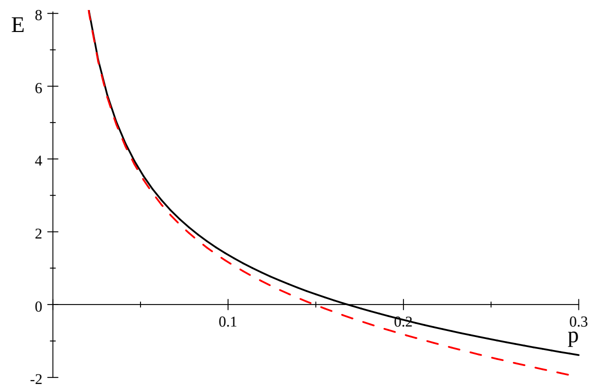

Figure C.1: The solid graph shows the excess demand for $\beta = 3$; here the equilibrium price is $p = 0.17$. The dashed graph shows excess demand for $\beta = 5$. Here the equilibrium price is $p = 0.15$.

shows the graphs of excess demand for $\beta = 3$ (solid) and for $\beta = 5$ (dashed). The higher value of β (associated with a larger supply at every price) leads to a lower equilibrium price.

The differential of E is

$$dE = \frac{\partial E}{\partial p}dp + \frac{\partial E}{\partial \beta}d\beta.$$

As we change β, the equilibrium price also changes. The equilibrium price causes excess demand to equal 0. Therefore, equilibrium requires that $dE = 0$:

$$dE = \frac{\partial E}{\partial p}dp + \frac{\partial E}{\partial \beta}d\beta = 0. \tag{C.1}$$

This equation states that an *exogenous* change in β, $d\beta$, induces an *endogenous* change in the price, dp, in order to maintain excess demand at 0. The partial derivatives of E are

$$\begin{aligned} \tfrac{\partial E}{\partial p} &= -0.6p^{-1.6} - \beta - 0.5p^{-0.5} < 0 \\ \tfrac{\partial E}{\partial \beta} &= -p < 0. \end{aligned} \tag{C.2}$$

The first line of equation C.2 says that an increase in p, at fixed β, decreases excess demand. The second inequality says that an increase in β, at fixed p, decreases excess demand. The endogenous price must adjust to a change in β in order to keep excess demand at 0, in order to satisfy our equilibrium condition. Inspection of these two partial derivatives tells us that if β increases (thereby lowering E), there must be an offsetting *decrease* in p, in order to maintain excess demand at 0.

Here we can figure out how p must change in response to a change in β simply by thinking a bit about the implication of the signs of the partial derivatives. Many problems are too complicated for that kind of casual reasoning to be useful. Therefore, we proceed systematically, using equations C.1 and C.2:

$$0 = \tfrac{\partial E}{\partial p}dp + \tfrac{\partial E}{\partial \beta}d\beta = \left(-0.6p^{-1.6} - \beta - 0.5p^{-0.5}\right)dp + (-p)\,d\beta \Rightarrow$$
$$\left(-0.6p^{-1.6} - \beta - 0.5p^{-0.5}\right)dp = p d\beta.$$

The first equality repeats equation C.1; the second uses the information in equation C.2. Rearranging this equation gives the implication in the second line. We solve this equation, dividing both sides by $d\beta$ and also dividing both sides by $\left(-0.6p^{-1.6} - \beta - 0.5p^{-0.5}\right)$ to write

$$\frac{dp}{d\beta} = \frac{p}{\left(-0.6p^{-1.6} - \beta - 0.5p^{-0.5}\right)}.$$

This equation is our comparative static expression. The numerator of the ratio on the right side is positive and the denominator is negative, so $\frac{dp}{d\beta} < 0$.

C.1 Comparative statics for the tea example

The equilibrium condition for the tea-in-China example is

$$\underbrace{20 - q^{\text{China}}}_{L\left(q^{\text{China}}\right)} \underset{=}{=} \underbrace{\frac{1}{1+b}\left(18 - \left[10 - q^{\text{China}}\right]\right)}_{R\left(q^{\text{China}},b\right).} . \tag{C.3}$$

In the text we solve this equation to obtain q^{China} as an explicit function of the model parameters. In cases where the equilibrium condition is too complicated to solve explicitly, we can still obtain information merely by using the equilibrium condition, which gives the endogenous variable as an implicit function of the exogenous parameters. The second line of equation C.3 shows that the left side is denoted as $L\left(q^{\text{China}}\right)$ and the right side as $R\left(q^{\text{China}}, b\right)$.

An exogenous change in b alters the right side without having a direct effect on the left side. In order for the equality $L\left(q^{\text{China}}\right) = R\left(q^{\text{China}}, b\right)$ to continue to hold after the change in b, there must be a compensating change in q^{China}. The change in R, denoted dR, must equal the change in L, denoted dL. These changes are called the "differentials" of R and L. Equilibrium requires

$$dL\left(q^{\text{China}}\right) = dR\left(q^{\text{China}}, b\right). \tag{C.4}$$

Using the definition of the differential and the fact that the right side of the equilibrium condition (R) depends on both b and q^{China}, and the left side (L) depends only on q^{China}, we have

$$\begin{aligned} dL &= \tfrac{\partial L\left(q^{\text{China}}\right)}{\partial q^{\text{China}}} dq^{\text{China}} \\ dR &= \tfrac{\partial R\left(q^{\text{China}}, b\right)}{\partial q^{\text{China}}} dq^{\text{China}} + \tfrac{\partial R\left(q^{\text{China}}, b\right)}{\partial b} db. \end{aligned}$$

Substituting these expressions into the definitions of the differentials, and using the equilibrium requirement $dL = dR$ gives

$$\frac{\partial L\left(q^{\text{China}}\right)}{\partial q^{\text{China}}} dq^{\text{China}} = \frac{\partial R\left(q^{\text{China}}, b\right)}{\partial q^{\text{China}}} dq^{\text{China}} + \frac{\partial R\left(q^{\text{China}}, b\right)}{\partial b} db.$$

Collecting terms gives

$$\left(\frac{\partial L\left(q^{\text{China}}\right)}{\partial q^{\text{China}}} - \frac{\partial R\left(q^{\text{China}}, b\right)}{\partial q^{\text{China}}}\right) dq^{\text{China}} = \frac{\partial R\left(q^{\text{China}}, b\right)}{\partial b} db.$$

Rearranging this equation gives

$$\frac{dq^{\text{China}}}{db} = \frac{\frac{\partial R\left(q^{\text{China}}, b\right)}{\partial b}}{\left(\frac{\partial L\left(q^{\text{China}}\right)}{\partial q^{\text{China}}} - \frac{\partial R\left(q^{\text{China}}, b\right)}{\partial q^{\text{China}}}\right)} > 0.$$

Using rules of differentiation, we have

$$\begin{aligned} &\tfrac{\partial L\left(q^{\text{China}}\right)}{\partial q^{\text{China}}} = -1 < 0, \; \tfrac{\partial R\left(q^{\text{China}}, b\right)}{\partial q^{\text{China}}} = \tfrac{1}{1+b} > 0 \\ \text{and } &\tfrac{\partial R\left(q^{\text{China}}, b\right)}{\partial b} = \tfrac{-1}{(1+b)^2}\left(18 - \left[10 - q^{\text{China}}\right]\right) < 0. \end{aligned}$$

Substituting the expressions for the partial derivatives into the previous equation and simplifying yields equation C.5

$$\frac{dq^{\text{China}}}{db} = \frac{q^{\text{China}} + 8}{b^2 + 3b + 2} > 0. \tag{C.5}$$

Equations 2.3 and C.5 both show how q^{China} responds to a change in transport costs, b. The right side of equation C.5 involves the unknown value q^{China}, whereas the right side of equation 2.3 involves only numbers and the exogenous parameter b. In this respect, the comparative statics expression 2.3 is more informative than equation C.5. The approach using differentials is useful when it is difficult to solve the equilibrium condition to obtain the explicit expression for the endogenous variable. Equation C.5 tells us only that an increase in b increases sales in China. Often we use models to obtain "qualitative" rather than "quantitative" information, e.g. we care more about the direction than the magnitude of the change.

C.2 Comparative statics for the two-period resource model

Following the procedure in the previous section, we denote the left side of the equilibrium condition, equation 3.5, as $L(y,\cdot) = (a - by - c)$ and the right side as $R(y,\cdot) = \rho(a - b(x-y) - c)$. The "$\cdot$" notation is shorthand for all of the exogenous variables, the parameters of the model: a, b, c, x, ρ. With this notation, we rewrite the equilibrium condition, equation 3.5 as $L(y,\cdot) = R(y,\cdot)$.

Equilibrium requires that a change in an exogenous parameter, such as the demand slope b, be offset by a change in the endogenous variable, period 0 supply, y: $dL = dR$. We totally differentiate the equilibrium condition, equation 3.5, with respect to y and b, to write

$$dL = \frac{\partial L}{\partial y}dy + \frac{\partial L}{\partial b}db = \frac{\partial R}{\partial y}dy + \frac{\partial R}{\partial b}db = dR.$$

Rearrange the differentials on the two sides of the equality to write

$$\left(\frac{\partial L}{\partial y} - \frac{\partial R}{\partial y}\right)dy = \left(\frac{\partial R}{\partial b} - \frac{\partial L}{\partial b}\right)db \Rightarrow \frac{dy}{db} = \frac{\frac{\partial R}{\partial b} - \frac{\partial L}{\partial b}}{\frac{\partial L}{\partial y} - \frac{\partial R}{\partial y}}.$$

To evaluate this expression we use

$$\frac{\partial L}{\partial y} = -b, \quad \frac{\partial R}{\partial y} = \rho b, \quad \frac{\partial L}{\partial b} = -y, \quad \frac{\partial R}{\partial b} = -\rho\left(x - y\right).$$

Putting these results together, we have

$$\frac{dy}{db} = \frac{\frac{\partial R}{\partial b} - \frac{\partial L}{\partial b}}{\frac{\partial L}{\partial y} - \frac{\partial R}{\partial y}} = \frac{-\rho\left(x-y\right) - \left(-y\right)}{-b - \rho b}$$

Simplifying the right side of this equation produces the comparative statics expression

$$\frac{dy}{db} = \frac{-\rho x + \left(1+\rho\right) y}{-b\left(1+\rho\right)}. \tag{C.6}$$

The denominator of the right side of this equation is negative, but without additional information we cannot determine whether the numerator is positive or negative. The most that we can say, using only the information in equation C.6, is that $\frac{dy}{db} < 0$ if and only if $-\rho x + (1+\rho)\, y > 0$, i.e. if and only if $y > \frac{\rho}{1+\rho} x$.

It is instructive to compare the derivatives $\frac{dy}{db}$ in equations 3.6 and C.6. Both of them are correct, but the former contains more information; it tells us that $\frac{dy}{db} < 0$, whereas equation C.6 gives us only a condition ($y > \frac{\rho}{1+\rho}x$) under which $\frac{dy}{db} < 0$. The fact that the two approaches yield different amounts of information is not surprising, because the first approach *begins with* more information: it uses the explicit expression for y as a function of model parameters. In contrast, the second approach uses only the equilibrium condition. More information is preferred to less, so in this sense the first approach is better than the second. But bear in mind that the first approach is not always available to us, because many models are too complicated to yield explicit solutions for the endogenous variables.

Appendix D

Comparison of monopoly and competitive equilibria

To emphasize that there is nothing peculiar about the possibility that the monopoly and competitive firm might choose the same level of sales (as in Chapter 3.2) , consider an even simpler, one-period model. In this model, the firm (either a monopoly or a representative competitive firm) can produce up to 10 units at constant costs 4. Production beyond that level is not feasible; equivalently, the marginal cost of production becomes infinite at $y = 10$. In the first scenario, the inverse demand function is $p = 30 - y$, and in the second scenario it is $p = 15 - y$. Figures D.1 and D.2 show the demand functions in these two cases (the solid lines), and the corresponding marginal revenue curves (the dashed lines). In both cases the marginal production cost is 4 for $y < 10$ and infinite for $y > 10$.

In the high-demand scenario (Figure D.1), the constraint $y \leq 10$ is binding for both the monopoly and the competitive firm, so both firms produce $y = 10$. In this case, the price is also the same under the competitive firm or the monopoly. In the low-demand scenario (Figure D.2), the constraint is binding for the competitive firm, which produces $y = 10$. Here, the marginal revenue curve (which lies below the demand curve) equals marginal cost at $y = 5.5$. The monopoly produces less than competitive firms, and receives a higher price.

There is nothing special about the possibility that a monopoly and a competitive firm might produce at the same level. The outcome depends on the relation between the point at which the cost function becomes vertical ($y = 10$ in this example) and the demand and marginal revenue functions.

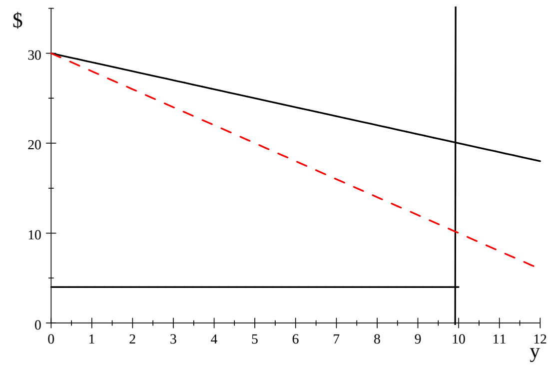

Figure D.1: The solid line shows the demand function $p = 30 - y$, and the dashed line is the marginal revenue function corresponding to this demand function. Marginal costs are constant at 4 for $y < 10$ and infinite for $y > 10$.

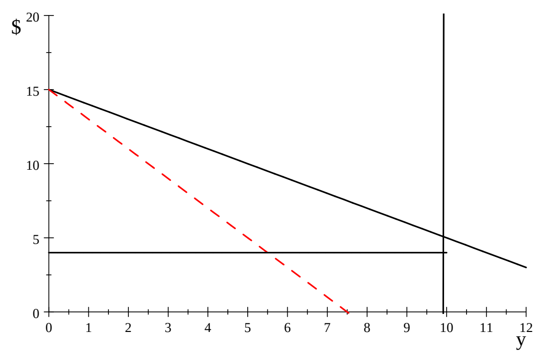

Figure D.2: The solid line shows the demand function $p = 15 - y$, and the dashed line is the marginal revenue function corresponding to this demand function. Marginal costs are constant at 4 for $y < 10$ and infinite for $y > 10$.

Appendix E

Derivation of the Hotelling equation

By definition, T is the last period during which extraction is positive, so extraction at $T+1$ is zero. In the class of problems we consider, extraction is also positive at earlier times: $y_t > 0$ for all $t < T$. This fact means that we can make small changes (perturbations) in any of the y_t's, and offsetting changes in other y_t's, without violating the non-negativity constraints on extraction, or on the stocks. A perturbation is "admissible" if it does not violate these constraints.

A "candidate" is a series of extraction and stock levels that satisfy the non-negativity constraints. At the optimum, any admissible perturbation of the candidate yields zero first order change in the payoff. In the two-period setting, only "one-step" perturbations, in which we make a small change in period-0 extraction and an offsetting change in period-1 extraction, are possible. In a multiperiod setting, in contrast, many types of perturbations are possible. For example, we can reduce extraction by ε in period t, make no change in period $t+1$, and increase extraction by $\varepsilon/3$ in each of the subsequent three periods. To test the optimality of a particular candidate, we have to be sure that no admissible perturbation, however complicated, creates a first order change in the payoff. With many possible perturbations, that sounds like a difficult job. However, the task turns out to be simple, because *any* admissible perturbation, no matter how complicated, can be broken down to a series of "one-step" perturbations.

Therefore, we can check whether a candidate is optimal by considering only the one-step perturbations affecting pairs of adjacent periods. Let t be any period less than T, so that extraction is positive in periods t and $t+1$. Because we are considering one-step perturbations that affect only these two periods, we only have to check that the perturbation has zero first order effect on the combined payoffs during these two periods. The combined payoff in these two periods, under the perturbation is

$$\begin{gathered} g\left(\varepsilon; y_t, x_t, y_{t+1}\right) = \rho^t[\left(p_t\left(y_t+\varepsilon\right) - c\left(x_t, y_t+\varepsilon\right)\right) \\ +\rho\left(p_{t+1}\left(y_{t+1}-\varepsilon\right) - c\left(x_{t+1}-\varepsilon, y_{t+1}-\varepsilon\right)\right)]. \end{gathered} \tag{E.1}$$

This gain function and the gain function from the two-period problem, equation 4.6, are the same, except for the time subscripts (and the fact that ρ^t multiplies the right side of equation E.1). In the two-period setting, we noted that an optimal candidate has to satisfy the first order condition 4.7. The necessary condition in the T-period setting is exactly the same, except for the time subscripts:

$$\frac{dg\left(\varepsilon; y_t, x_{t+1}, y_{t+1}\right)}{d\varepsilon}_{|\varepsilon=0} = 0.$$

Evaluating this derivative (repeating the steps in Box 4.2) produces the Euler equation 5.2.

Appendix F

Algebra of taxes

This appendix collects technical details for the chapter on taxes.

F.1 Algebraic verification of tax equivalence

Denote the producer price as p^s (for supply) and the consumer price as p^c (for consumption) and write the "market price" as p. If consumers pay the tax, the prices are $p^s = p$ and $p^c = p^s + \nu$ (producers receive the market price and consumers pay this price plus the tax). If producers pay the tax, $p^s = p - \nu$ and $p^c = p$ (consumers pay the market price and producers receive this price minus the tax). We want to confirm that tax-inclusive prices are the same regardless of who directly pays the tax.

If consumers pay the tax, the supply equal demand condition is

$$S(p) = D(p + \nu). \tag{F.1}$$

Let $p^*(\nu)$ be the (unique) price that solves this equation; this is the equilibrium producer price (a function of ν) when consumers pay the tax: $p^*(0)$ is the equilibrium price when $\nu = 0$. Because consumers (directly) pay the tax, the price producers receive (the "supply price") equals $p^*(\nu)$ and the price consumers pay equals $p^*(\nu) + \nu$.

If, instead, producers directly pay the tax, the equilibrium condition is

$$S(p - \nu) = D(p). \tag{F.2}$$

Substitute $p = p^* + \nu$ into this equation to write equation $(F.2)$ as

$$S(p^*) = D(p^* + \nu).$$

The last equation reproduces equation $(F.1)$ evaluated at $p = p^*(\nu)$, the unique solution to that equation. Thus, the two equations $(F.1)$ and $(F.2)$ lead to the same producer and consumer prices.

F.2 The open economy

For a closed economy, domestic supply equals domestic demand: there is no trade. For an open economy, the difference between domestic demand and supply equals the amount imported or exported. Tax equivalence holds in a closed economy, where all sources of supply or demand are subject to the tax, but not in an open economy. In Chapter 10.1 we noted that in a closed economy, the "Polluter Pays Principle" may be vacuous, because (under some conditions) the tax equivalence result implies that it does not matter whether the polluter or the pollutee pays the environmental tax. Because tax incidence does not hold in the open economy, it *does* matter whether a consumer or producer tax is used.

We use an example to compare tax incidence in a closed and an open economy. First consider the case where the economy is closed. Suppose that domestic demand is $q^d = 10 - p$, domestic supply equals $q^s = bp$, and foreign supply is $q^{s,for} = cp$. Column 2 of Table F.1 shows that, for the closed economy, the consumer and producer tax incidence does not depend on which agent, consumers or producers, directly pays the tax. The incidences in this column are calculated using the following steps:

1. Calculate the equilibrium price in the absence of tax by setting the untaxed supply equal to the untaxed demand.

2. Calculate the equilibrium consumer price and producer price when one of these agents directly pays the tax, by setting the (taxed) demand equal to the (taxed) supply.

3. Use the tax-inclusive consumer and producer price for the two cases (where one agent or the other directly pays the tax) and the zero-tax price to calculate the incidences.

The third column of the table shows that in the open economy, the incidences do depend on which (domestic) agent directly pays the tax. For example, to calculate the incidences in the open economy when consumers directly pay the tax (regardless of the source of supply), we use essentially the same steps as above. The market clearing condition in the absence of a tax is $10 - p = (b + c)\,p$. We solve this to find the zero-tax price. If consumers pay the tax, the market clearing condition is $10 - (p + \tau) = bp + cp$, where now we understand that p is the price received by both domestic and foreign firms, and $p + t$ is the consumer tax-inclusive price. We solve this equation to find the equilibrium producer and consumer prices. Using the formula for tax incidence, we obtain the expressions in the third row and third column of Table F.1. An exercise asks readers to use this procedure to derive the formulae in the table.

	closed economy $q^d = 10 - p$ $q^s = bp$ market clearing $10 - p = bp$	open economy $q^d = 10 - p$ $q^s = bp$ and $q^{s,for} = cp$ market clearing $10 - p = (b + c)\,p$
consumers pay tax	consumer incidence $\frac{b}{1+b}100\%$ producer incidence $\frac{1}{1+b}100\%$	consumer incidence $\frac{b+c}{1+b+c}100\%$ producer incidence $\frac{1}{1+b+c}100\%$
domestic producers pay tax	consumer incidence $\frac{b}{1+b}100\%$ producer incidence $\frac{1}{1+b}100\%$	consumer incidence $\frac{b}{1+b+c}100\%$ producer incidence $\frac{1+c}{1+b+c}100\%$

Table F.1 consumer and producer tax incidence in closed and open economy

In an open economy, domestic supply does not equal to domestic demand. Taxing consumers causes the market demand function to shift in, lowering the price that both domestic and foreign producers face and increasing the consumer's tax-inclusive price. Taxing only domestic supply causes the domestic supply function to shift in, increasing the consumer price, decreasing the domestic tax-inclusive price, and shifting supply from domestic to foreign producers. Under the consumer tax, both the domestic and foreign producers receive the same price. Under the (domestic) producer tax, consumers and foreign producers face the same price, and domestic producers receive a lower after-tax price.

This example shows that although in a closed economy producer and consumer taxes are equivalent, the two taxes are not equivalent in an open economy. The rest of this appendix considers only the closed economy.

Exercise: Derive the tax incidences shown in Table F.1.

F.3 Approximating tax incidence

In the closed economy, it does not matter whether consumers or producers are charged the tax. Suppose that consumers are charged the tax, so that the equilibrium condition is equation F.1. This equation expresses the equilibrium price as an implicit function of the tax: as the tax changes, the equilibrium price p changes. The consumer incidence (expressed as a fraction instead of a percent) equals

$$\frac{p^*(\nu) + \nu - p^*(0)}{\nu} = \frac{p^*(\nu) - p^*(0)}{\nu - 0} + 1 = \frac{\Delta p}{\Delta \nu} + 1. \tag{F.3}$$

The numerator on the left side equals the change in price that consumers pay. We obtain the first equality by simplifying, i.e. using the fact that $\frac{\nu}{\nu} = 1$, and subtracting 0 from the denominator. We subtract 0 in order to emphasize that both the numerator and the denominator are changes: the numerator is the change in price, in moving from a 0 tax to a non-zero tax, and the denominator is the change in the tax, $\nu - 0$. We obtain the second equality by using the "delta notation": Δ means "change in". The next step requires a formula for an approximation of $\frac{\Delta p}{\Delta \nu}$, which we obtain using the fact that the derivative $\frac{dp}{d\nu}$ is approximately equal to $\frac{\Delta p}{\Delta \nu}$.

Treating $p = p(\nu)$ (i.e. price as a function of the tax – and dropping the "*" to simplify notation) we can differentiate both sides of the equilibrium condition F.1 to write

$$\frac{dS(p)}{dp}\frac{dp}{d\nu} = \frac{dD(p+\nu)}{dp^c}\left(\frac{dp}{d\nu} + 1\right).$$

Divide both sides by the equilibrium quantity, using $S = D$, and multiply by the equilibrium price p to write

$$\frac{dS(p)}{dp}\frac{p}{S}\frac{dp}{d\nu} = \frac{dD(p+\nu)}{dp^c}\frac{p}{D}\left(\frac{dp}{d\nu} + 1\right). \tag{F.4}$$

Because we are considering an approximation for small ν, we evaluate equation F.4 at $\nu = 0$. Using the definitions in equation 10.1, and evaluating equation F.4 at $\nu = 0$, we rewrite that equation as

$$\theta \frac{dp}{d\nu} = -\eta \left(\frac{dp}{d\nu} + 1 \right).$$

We can solve this equation for $\frac{dp}{d\nu}$ to obtain

$$\frac{dp}{d\nu} = -\frac{\eta}{\theta + \eta}. \tag{F.5}$$

This equation shows the derivative of the equilibrium price with respect to the tax, evaluated at a 0 tax. Notice that $\frac{dp}{d\nu} < 0$: the tax, although paid by consumers, reduces the equilibrium price that producers receive.

We use the fact that

$$\frac{\Delta p}{\Delta \nu} \approx \frac{dp}{d\nu}$$

and equations F.3 and F.5 to write the expression for the consumer incidence as

$$\frac{\Delta p}{\Delta \nu} + 1 \approx \frac{dp}{d\nu} + 1 = 1 - \frac{\eta}{\theta + \eta} = \frac{\theta}{\theta + \eta}.$$

The tax incidence for producers equals

$$\frac{\text{reduction in producer price}}{\text{level of (unit) tax}}.$$

Initially the tax is 0, so the level of the tax (once it is imposed) is $\nu - 0 = \Delta \nu$. The producer tax incidence is

$$\text{producer incidence: } \frac{-\Delta p}{\Delta \nu} \approx \frac{\eta}{\theta + \eta}.$$

This expression involves $-\Delta p$ rather than Δp because the definition of the producer incidence involves the "price reduction", not the "price change". If the price change is, for example, -3, then the reduction is 3.

Exercise: Suppose that consumers are charged the tax, as above. Let the demand function be $D(p) = p^{-\eta}$ with $\eta > 1$ and suppose that firms have constant marginal cost, c. Evaluate the consumer and producer tax incidence under the monopoly, as a function of η. Compare with the consumer and producer tax incidence under competition, with the same demand and cost functions. *Hint* Mimic the derivation above, replacing marginal revenue with price.

F.4 Approximating deadweight loss

The graphical representation of the deadweight cost of the tax is the area of the triangle in Figure 10.1. To verify equation 10.4 for the case of linear supply and demand we use the formula for the area of a triangle: one half base times height. Turn the triangle *bcd* in Figure 10.1 "on its side", so that the base of the rotated triangle is *bd*, and split the triangle into two triangles, *bcg* and *dcg*. The area of *bcd* equals the sum of the area of the two smaller triangles. Denote the consumer tax incidence (as a fraction, not a percent) as $1-\phi$, so the producer incidence is ϕ. The length of *bd* is ν, the tax, so the length of the base of *gb* is $\phi\nu$. Denote the absolute value of the slope of *bc* as S_1 and denote the slope of *dc* as S_2. Using the formula "slope = rise/run", $S_1 = \frac{\phi\nu}{\Delta q}$, or $\Delta q = \frac{\phi\nu}{S_1}$. Therefore, the area of triangle *bcg* is $\frac{1}{2}\phi\nu\frac{\phi\nu}{S_1} = \frac{1}{2}\frac{\phi^2}{S_1}\nu^2$. Using the same reasoning, the area of the triangle *dcg* is $\frac{1}{2}\frac{(1-\phi)^2}{1-S_2}\nu^2$. The sum of the areas is $\frac{1}{2}\left(\frac{\phi^2}{S_1} + \frac{(1-\phi)^2}{1-S_2}\right)\nu^2$, i.e. it is proportional to the square of ν.

To approximate the DWL when the supply and demand functions are not linear, we again begin with the formula for the area of a triangle, turned on its side. The base of the triangle is the tax, ν. Denote the height of this triangle as Δq, the change in quantity demanded. We have (by multiplying and dividing)

$$\Delta q = \frac{\Delta q}{\Delta p}\Delta p = \frac{\Delta q}{\Delta p}\left(\frac{\Delta p}{\Delta \nu}\right)\Delta\nu = \left(\frac{\Delta q}{\Delta p}\frac{p}{q}\right)\frac{q}{p}\left(\frac{\Delta p}{\Delta \nu}\right)\Delta\nu. \tag{F.6}$$

Equation 10.2 and the definition of the supply elasticity imply, respectively, the following two equations

$$\frac{\Delta p}{\Delta \nu} \approx \frac{dp}{d\nu} = -\frac{\eta}{\theta+\eta} \text{ and } \left(\frac{\Delta q}{\Delta p}\frac{p}{q}\right) \approx \theta.$$

Inserting these formulae into equation F.6 gives the approximation

$$\Delta q \approx \frac{\theta\eta}{\theta+\eta}\frac{q}{p}\nu. \tag{F.7}$$

Here we used the fact that $\Delta\nu = \nu - 0 = \nu$, because we are taking the approximation in the neighborhood of a zero tax. This result and the formula for the area of a triangle produces equation 10.4.

F.5 Cap and trade

This appendix provides more detail on the comparison of taxes and cap and trade. We explain how a cap and trade system works, and why the equilibrium level of each firm's emissions does not depend on whether firms are given permits or have to buy them. We then explain the sense in which a cap and trade policy is equivalent to an emissions tax. We use that equivalence to approximate the fraction of permits that firms would have to be given, in order to make them just as well off under cap and trade as they are in the absence of regulation.

The basic ingredients of cap and trade. The regulator chooses the cap on emissions, denoted Z. The many competitive firms are able to buy and sell permits. This buying and selling is the "trade" part of the cap and trade policy. Each of these firms takes the price of an emissions permit as given. Denote the equilibrium price of permits as $p^e(Z)$. This relation recognizes that (as in all markets) the equilibrium price depends on the supply. Here, the supply is a number, Z. Due to the (assumed) fixed relation between output and emissions, by choice of units we can set one unit of output to equal one unit of emissions.

Claim #1: The permit price and firm-level emissions are independent of the allocation of permits The equilibrium permit price depends on the aggregate number of permits, Z. However, if firms are price taking and profit maximizing, and if the permit market works well, then firm-level pollution levels are *independent* of the distribution of allowances, e.g. whether firms are given or sold the permits. To verify this claim, we show that each firm's demand for permits is independent of its own allocation. Consider an arbitrary firm that is given an allowance A (possibly equal to zero). This price-taking faces the output price, p, and the permit price, p^e, and wants to maximize profits:

$$\underline{pq - c(q)} + \underbrace{p^e(A - q)}.$$

The underlined term equals the firm's revenue from selling the good minus its cost of production; the under-bracketed term equals the firm's profits from selling (if $A > q$) or its costs of buying (if $A < q$) permits.

The first order condition to the firm's problem states that price equals marginal cost. Marginal cost here equals the sum of the "usual" marginal cost ($\frac{dc}{dq}$), and the cost of buying an emissions permit, p^e. The first order condition

$$p = \frac{dc}{dq} + p^e, \tag{F.8}$$

does not depend on its permit allocation, A. A firm's decision about how much to produce, and thus about how many permits to use, does not depend on the allocation of permits.

A firm that buys permits has to pay p^e for the additional permit needed to produce an additional unit. A firm that sells permits incurs an *opportunity cost* p^e in using an additional permit: by using that permit it is no longer able to sell it. Thus, regardless of whether the firm is a net buyer or seller of permits, it incurs the cost p^e of using an additional unit. Recent research finds empirical support for Claim 1 (Box 10.1).

Claim #2: There exists a quota-equivalent emissions tax To simplify the exposition, we assume that all firms have the same cost function, so that we can use the representative firm model. As in Chapter 2.4, we denote the cost function for the representative firm as $c(Q)$. Using the fact that $Q = Z$ (because one unit of output produces one unit of emissions), equation F.8 implies

$$\begin{gathered} p(Q)_{|Q=Z} = \frac{dC(Q)}{dQ}_{|Q=Z} + p^e(Z) \quad \text{or} \\ \left(p(Q) - \frac{dC(Q)}{dQ}\right)_{|Q=Z} = p^e(Z). \end{gathered} \tag{F.9}$$

The second equation shows that the equilibrium permit price equals the difference between the inverse demand function, $p(Q)$, and the marginal cost function, $\frac{dC(Q)}{dQ}$.

Figure F.1 shows linear (product) demand and marginal cost curves, and the equilibrium permit price (dashed line) as the vertical difference between the two (equal to the left side of equation F.9). The dashed curve is the inverse demand for pollution permits. For this example, the equilibrium quantity (= emissions) absent regulation is 3.33. If the regulator chooses $Z \geq 3.33$, then the regulation is vacuous, and the permit price is zero. But for $Z < 3.33$, the emissions constraint is binding, and the permit price is positive. Every value of Z below the unregulated "Business as usual" level (3.33) corresponds to a different equilibrium permit price.

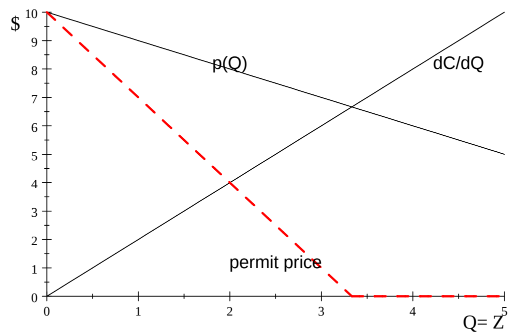

Figure F.1: Solid lines: inverse demand and marginal cost. Dashed curve: the equilibrium permit price, p^e, is the vertical difference between inverse demand and marginal cost.

If the representative firm faces a tax ν, the equilibrium condition (price equals "usual" marginal cost plus the tax) is

$$p(Z) = \frac{dC(Q)}{dQ} + \nu. \tag{F.10}$$

Comparing equation F.10 to the first line of equation F.9 shows that the tax $\nu = p^e(Z)$ induces the competitive industry to produce at the same level as under the cap and trade policy with cap Z.

Claim #3: There is a simple formula for compensating firms A cap and trade policy with cap Z determines the amount of emissions. We showed that there is an equivalent tax that leads to the same amount of emissions. Firms have the same level of producer surplus if they face the cap Z and all permits are auctioned (i.e., there is no grandfathering), or if they face the tax that "supports" the level of pollution Z. Under a cap and trade policy the regulator can reduce the cost to the firms by giving them (grandfathering) some permits, instead of auctioning all of them. (Under the equivalent tax, the regulator can compensate firms by giving them some of the tax revenue.) What fraction of permits would the regulator have to give firms, to make them (almost) as well off under regulation as under Business as Usual?

This question has a simple answer in our setting. If the regulator auctions all of the permits (gives none to the firms) then from the standpoint of firms, it is exactly as if they face the tax $\nu = p^e(Z)$. Figure 10.1 shows that the firms' loss in surplus, due to a tax ν (or to being forced to buy all of its emissions permits at the price $\hat{p}(Z) = \nu$) is the area $fcba = fgba + bgc$. Denote the producer incidence under the tax (a fraction) as ϕ. If firms are given (instead of being forced to buy) the fraction $\frac{\eta}{\theta+\eta}$ of permits, then the value of this gift is ϕ times the potential tax revenue. (Review equation 10.3.) This value equals the area of the rectangle $fgba$. Firms' net loss equals their loss in producer surplus minus the value of the gift, the area of the triangle bcg. This triangle is the small correction that is needed to make firms whole. The fraction ϕ is (typically) much less than 1, so even with the correction, it would be necessary to give firms only a fraction of the permits, to compensate them for the regulation.

Appendix G

Continuous time

Consider the growth equation with a harvest rule $y(x)$; this rule determines the level of harvest as a function of the stock, x. Equation 13.3 shows the discrete time dynamics for two particular harvest rules. Later we encounter other harvest rules, so here we use the general formulation $y(x)$. With this harvest rule, the next-period and current-period stocks are related according to

$$x_{t+1} - x_t = F(x_t) - y(x_t) = [F(x_t) - y(x_t)]\,1. \tag{G.1}$$

Multiplying $F(x_t) - y(x_t)$ by 1, as in the last equality, obviously does not change the quantity.

We have to measure time in specific units. For example, it is meaningful to say "That was three years ago," but we would never say "That was three ago." We choose the unit of time to equal one year; this choice is arbitrary: we could have chosen a unit to equal one second or one century. There is no reason (apart from convenience) to assume that the length of a period equals one unit of time.

We use the symbol Δ to represent the length of a period. Given that our unit of time is a year, the symbol $\Delta = 10$ means that a period lasts for a decade. If a period lasts for a day, then $\Delta = \frac{1}{364}$. In order for our model to show explicitly the length of a period, we can replace the number 1 wherever it appears in equation G.1 (including in the subscripts) with Δ; the equation becomes

$$\begin{gathered} x_{t+\Delta} - x_t = [F(x_t;\Delta) - y(x_t;\Delta)]\,\Delta \Rightarrow \\ \tfrac{x_{t+\Delta}-x_t}{\Delta} = F(x_t;\Delta) - y(x_t;\Delta)\,. \end{gathered} \tag{G.2}$$

By introducing the parameter Δ, we have made a subtle change in the definition of $F(x_t)$ and $y(x_t)$; these are now *rates*, i.e. they give growth and harvest *per unit of time* (one year). To take into account this change, we replace $F(x_t)$ and $y(x_t)$ with $F(x_t;\Delta)$ and $y(x_t;\Delta)$. If $\Delta = \frac{1}{364}$ and if the growth per year is 0.8, and harvest per year is 0.2 , then the amount of growth and harvest over one period (one day, for $\Delta = \frac{1}{364}$) equals $\frac{0.8}{364}$ and $\frac{0.2}{364}$, respectively. (The change over one day is $x_{t+\Delta} - x_t = (0.8 - 0.2)\,\Delta = (0.8 - 0.2)\,\frac{1}{364}$.)

Now that we explicitly recognize that growth and harvest are rates, we no longer need to require that $y \leq x$. For example, suppose that $x = 40$ and $y = 60$. It is not possible to extract 60 units of biomass if the stock of biomass equals only 40. However, it is certainly possible to harvest at an annual rate of 60 for a short period of time. If $\Delta = \frac{1}{364}$ and $y = 60$, then after 10 periods (= 10 days) we have extracted $\frac{60}{364} \times 10 = 1.65$ units of biomass. In general, if Δ is sufficiently small, then the annual harvest rate y can be arbitrarily large without violating the non-negativity constraint on the stock of fish.

The last line of equation G.2 shows the ratio $\frac{x_{t+\Delta}-x_t}{\Delta}$, equal to the change in stock per change in time. With $\Delta = \frac{1}{364}$, this ratio is the change in the stock per day. As $\Delta \to 0$, the ratio $\frac{x_{t+\Delta}-x_t}{\Delta}$ converges to a time derivative. We define

$$F(x) = \lim_{\Delta \to 0} F(x;\Delta) \text{ and } y(x) = \lim_{\Delta \to 0} y(x;\Delta).$$

With this definition, the continuous time limit of the last line of equation G.2 is

$$\frac{dx_t}{dt} = F(x_t) - y(x_t). \tag{G.3}$$

Equation G.2 is a difference equation, and equation G.3 is a differential equation. They both describe how x changes over time. *When studying stability we use the continuous time model.*

It is important to be clear about the relation between equations G.2 and G.3. By construction, they have the same steady states. In other respects, however, they may contain very different information. For example, suppose that we have two fish stocks; the first grows according to equation G.2 and the second grows according to equation G.3. We start both stocks at the same level, and let each evolve in the manner described by its equation of motion. Would these two stocks evolve in the same way, i.e. would the time-graphs of their trajectories look similar? In general, the answer is "no". If we want to change the length of a period (e.g. from $\Delta = 1$ to $\Delta = \frac{1}{10{,}000{,}000{,}000}$), while keeping the trajectory qualitatively unchanged, we have to re-calibrate the functions $F(x_t)$ and $y(x_t)$. However, if Δ is sufficiently small, then trajectories arising from the continuous and discrete time models are qualitatively similar, at least in the neighborhood of a steady state.

Appendix H

Bioeconomic equilibrium

Here we offer a slightly different perspective on the open access steady state. The zero profit condition, and the production function in equation 14.2, imply

$$0 = (pqx - w) E \Rightarrow x = \frac{w}{pq} = \frac{C}{p}, \tag{H.1}$$

where the last equality uses the definition $\frac{w}{q} = C$. The production function 14.2 and the steady state condition under logistic growth (harvest equal growth) give

$$y = qEx = \gamma x \left(1 - \frac{x}{K}\right). \tag{H.2}$$

Substituting equation H.1 into equation H.2 gives the *steady state* supply function

$$y = \gamma \frac{C}{p} \left(1 - \frac{C}{Kp}\right). \tag{H.3}$$

The steady state supply function for harvest gives the harvest level, as a function of the price, that is consistent with a steady state stock of the fish and zero profits in the fishery. Figure H.1 shows the supply function for parameter values $K = 50, \gamma = 0.03$, and $C = 5$. The notable feature is that this supply function bends backwards. For prices $p < \frac{1}{Kq} = 0.2$, supply increases with price, and is very price-elastic (flat). At higher prices, equilibrium supply decreases with the higher price. At a given stock, the higher price induces greater harvest, but the higher harvest reduces the steady state stock. The net effect in the steady state is that a higher price reduces equilibrium supply over the backward bending part of the curve.

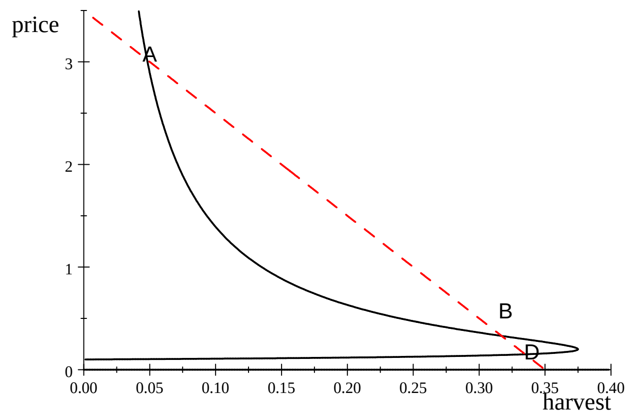

Figure H.1: The backward bending steady state supply function (solid) and a linear demand curve.

The dashed curve in the figure shows the linear demand curve, $p = 3.5 - 10y$. There are three "bioeconomic equilibria", combinations of output and price where supply equals demand and the stock is in a steady state. The equilibria A and D, corresponding to high price and low harvest, and low price and high harvest, are stable; the intermediate equilibrium is unstable, just as we saw in Section 14.1.2.

In order to examine the stability of the different steady states, we introduce a fictitious "Walrasian auctioneer". This auctioneer calls out an arbitrary price. If, at that price, supply equals demand, the auctioneer has found an equilibrium. However, if at the price the auctioneer has called out, demand exceeds supply, then the auctioneer raises the price, in an effort to bring supply and demand into equilibrium.

Suppose that this auctioneer calls out a price slightly above the p coordinate of point B; at this slightly higher price, demand exceeds supply. In an effort to balance supply and demand, the auctioneer increases the price. The higher price initially elicits greater supply, but that reduces future stock, creating an even larger divergence between steady state supply and demand. The auctioneer continues to raise the price, toward the p coordinate of point A, where at last steady state supply equals demand. Thus, a price that begins slightly above the p coordinate of point B moves away from that point, so this price is unstable. Parallel arguments show that the prices corresponding to points A and D are stable steady states.

Appendix I

The Euler equation for the sole owner fishery

We find the optimality condition where both the stock and the harvest are strictly positive, i.e. along an interior optimum. As in Chapter 5, we determine the optimality condition using the perturbation method. Natural growth, the function $F(x_t)$, complicates the problem, but the logic is the same. We begin with a feasible "candidate trajectory" of harvest, y_0, y_1, y_2..., and the corresponding stock sequence, x_0, x_1, x_2... . We obtain the condition that must be satisfied if no perturbation increases the present discounted value of the payoff. As before, we consider a particular one-step perturbation: one that increases harvest in an arbitrary period (t) by ε, and makes an offsetting change in harvest in the next period $(t+1)$ in order to keep unchanged the stock in the subsequent period $(t+2)$. We can build a more complicated perturbation from a series of these one-step perturbations, but for the purpose of obtaining the necessary conditions, it suffices to consider the one-step perturbation.

We begin by finding the offsetting change needed in period $t+1$, in order to keep unchanged (relative to the unperturbed candidate) the stock in period $t+2$. If we increase harvest in period t by ε, the stock in period $t+1$ is

$$x_{t+1} = x_t + F(x_t) - (y_t + \varepsilon).$$

This relation implies

$$\frac{dx_{t+1}}{d\varepsilon} = -1. \tag{I.1}$$

The increased harvest in period t reduces the stock in period $t+1$ from x_{t+1} (the level under the candidate trajectory) to $x_{t+1} - \varepsilon$. Under the candidate trajectory, we plan to harvest y_{t+1} in period $t+1$. The offsetting change in y_{t+1}, required by the fact that we increased y_t by ε, and by our insistence that x_{t+2} be unchanged, is $\delta(\varepsilon)$. The notation $\delta(\varepsilon)$ emphasizes that δ, the change in y_{t+1}, depends on, ε, the change in y_t. Using the growth function, we have

$$x_{t+2} = (x_{t+1} - \varepsilon) + F(x_{t+1} - \varepsilon) - (y_{t+1} + \delta(\varepsilon)). \tag{I.2}$$

We require that the total change in x_{t+2} – including the changes in both periods t and $t+1$, be zero, i.e.,

$$\frac{dx_{t+2}}{d\varepsilon} = 0.$$

Using this condition, and differentiating both sides of equation I.2 implies

$$\begin{gathered} \frac{dx_{t+2}}{d\varepsilon} = 0 = -1 + \frac{dF(x_{t+1})}{dx_{t+1}}\frac{dx_{t+1}}{d\varepsilon} - \frac{d\delta}{d\varepsilon} \Rightarrow -1 - \frac{dF(x_{t+1})}{dx_{t+1}} - \frac{d\delta}{d\varepsilon} = 0 \Rightarrow \\ \frac{d\delta}{d\varepsilon} = -\left(1 + \frac{dF(x_{t+1})}{dx_{t+1}}\right). \end{gathered} \tag{I.3}$$

The first line differentiates both sides of equation I.2 with respect to ε, using the chain rule. We use equation I.1 to eliminate $\frac{dx_{t+1}}{d\varepsilon}$ to obtain the equation after the first "$\Rightarrow$", and rearrange that equation to obtain the second line.

The last line of equation I.3 provides the first piece of information: the required reduction in y_{t+1}, given that we increase y_t by ε, and given that we want to keep x_{t+2} unchanged. A one unit increase in y_t leads to a one unit direct reduction in x_{t+1} and $\frac{dF(x_{t+1})}{dx_{t+1}}$ units loss in growth; the loss in growth affects x_{t+2}. Therefore, if we increase y_t by ε units, we must decrease y_{t+1} by $\left(1 + \frac{dF(x_{t+1})}{dx_{t+1}}\right)\varepsilon$ units, to offset both the direct effect on x_{t+2} and the indirect effect that occurs via the reduced growth.

Under the perturbation, periods' t and $t+1$ contribution to the total payoff is ρ^t times

$$g(\varepsilon) = \left(p_t - \frac{C}{x_t}\right)(y_t + \varepsilon) + \rho\left(p_{t+1} - \frac{C}{x_{t+1} - \varepsilon}\right)(y_{t+1} + \delta).$$

If the candidate is optimal, then a perturbation must lead to a zero first order change in the gain function. Using the product and the quotient rules, we have

$$\begin{gathered}
\frac{dg(\varepsilon)}{d\varepsilon}_{|\varepsilon=0} = p_t - \frac{C}{x_t} + \rho\left[-\frac{C}{x_{t+1}^2}y_{t+1} + \left(p_{t+1} - \frac{C}{x_{t+1}}\right)\frac{d\delta}{d\varepsilon}\right] = \\
p_t - \frac{C}{x_t} + \rho\left[-\frac{C}{x_{t+1}^2}y_{t+1} - \left(p_{t+1} - \frac{C}{x_{t+1}}\right)\left(1 + \frac{dF(x_{t+1})}{dx_{t+1}}\right)\right] = 0 \Rightarrow \\
p_t - \frac{C}{x_t} = \rho\left[\left(p_{t+1} - \frac{C}{x_{t+1}}\right)\left(1 + \frac{dF(x_{t+1})}{dx_{t+1}}\right) + \frac{C}{x_{t+1}^2}y_{t+1}\right]
\end{gathered}$$

The last line is the Euler Equation 15.4.

Appendix J

Dynamics of the sole owner fishery

This appendix derives the continuous time analog of the Euler equation, and then derives the differential equation for harvest.

J.1 Derivation of equation 16.2

We could have begun with a continuous time problem, and derived equation 16.2 directly. That approach is mathematically preferable, but it requires methods that we have not discussed. Therefore, we proceed mechanically, taking equation 15.7 as our starting point, and showing how to manipulate it to produce the continuous time analog, equation 16.2.

We need to have in mind a unit of time. Because we want the discrete time and the continuous time models to be "close to each other", the unit of time should be small. As in Chapter 13.3, we begin with the model in which one period equals one unit of time, and then divide that period into smaller subperiods. Mechanically, we do this by replacing the number 1, the length of a period, with Δ. We also need to rewrite the discount factor as $\rho = \frac{1}{1+\Delta r}$ instead of $\rho = \frac{1}{1+r}$ Using the growth equation

$$x_{t+\Delta} - x_t = \left[F\left(x_t\right) - y\left(x_t\right)\right]\Delta,$$

we replace $F\left(x_t\right)$ and $y\left(x_t\right)$ with $F\left(x_t\right)\Delta$ and $y\left(x_t\right)\Delta$. Thus, the terms

$$\frac{dF\left(x_{t+1}\right)}{dx_{t+1}} \quad \text{and} \quad -\frac{C}{x_{t+1}^2}y_{t+1}$$

in equation 15.7 become, respectively,

$$\frac{dF\left(x_{t+\Delta}\right)}{dx_{t+\Delta}}\Delta \text{ and } -\frac{C}{x_{t+\Delta}^2}y_{t+\Delta}\Delta.$$

These substitutions mean that instead of having the length of a period be one unit of time (e.g. one minute), we now have the length of a period be Δ units of time. With these substitutions, we rewrite equation 15.7 as

$$R_t = \frac{1}{1+\Delta r}\left[R_{t+\Delta}\left(1+\frac{dF\left(x_{t+\Delta}\right)}{dx_{t+\Delta}}\Delta\right)+\frac{C}{x_{t+\Delta}^2}y_{t+\Delta}\Delta\right]. \tag{J.1}$$

Subtract $R_{t+\Delta}$ from both sides of equation J.1 and collect terms on the right side to rewrite the result as

$$R_t - R_{t+\Delta} = \left(\frac{1}{1+\Delta r}-1\right)R_{t+\Delta}+\frac{1}{1+\Delta r}\left[R_{t+\Delta}\left(\frac{dF\left(x_{t+\Delta}\right)}{dx_{t+\Delta}}\right)+\frac{C}{x_{t+\Delta}^2}y_{t+\Delta}\right]\Delta.$$

Divide both sides of this equation by Δ to write

$$-\frac{\left(R_{t+\Delta}-R_t\right)}{\Delta} = \frac{\left(\frac{1}{1+\Delta r}-1\right)}{\Delta}R_{t+\Delta}+\frac{1}{1+\Delta r}\left[R_{t+\Delta}\left(\frac{dF\left(x_{t+\Delta}\right)}{dx_{t+\Delta}}\right)+\frac{C}{x_{t+\Delta}^2}y_{t+\Delta}\right].$$

Now take the limit of both sides of this equation as $\Delta \to 0$, using

$$\lim_{\Delta\to 0}\frac{\left(R_{t+\Delta}-R_t\right)}{\Delta} = \frac{dR_t}{dt} \text{ and } \lim_{\Delta\to 0}\frac{\left(\frac{1}{1+\Delta r}-1\right)}{\Delta} = \lim_{\Delta\to 0}\frac{\left(\frac{1-1-\Delta r}{1+\Delta r}\right)}{\Delta} = -r$$

to write

$$-\frac{dR_t}{dt} = -R_t\left(r-\frac{dF\left(x_t\right)}{dx_t}\right)+\frac{C}{x_t^2}y_t.$$

Multiplying through by -1 gives equation 16.2.

J.2 The differential equation for harvest

Chapter 16.3.2 uses the differential equation for the sole owner harvest, $\frac{dy}{dt} = H\left(x,y\right)$. This appendix explains how we obtain this equation, i.e., how we obtain the function $H\left(x,y\right)$. The procedure uses the equations of motion for the stock, the rent, and the definition of rent. We repeat these three equations:

$$\frac{dx}{dt} = F(x) - y$$
$$\frac{dR_t}{dt} = R_t\left(r - \frac{dF(x_t)}{dx_t}\right) - \frac{C}{x_t^2}y_t \tag{J.2}$$
$$R_t = p(y_t) - \frac{C}{x_t}$$

The first equation is merely the constraint of the problem, i.e. it is "data" (given to us). The second equation is the Euler equation, expressed in terms of rent. Both of these equations are the continuous time versions of the discrete time model. The third equation is the definition of rent.

Because the third equation holds identically with respect to time (i.e., it holds at every instant of time), we can differentiate it with respect to time to write

$$\frac{dR_t}{dt} = \frac{d\left[p(y_t) - \frac{C}{x_t}\right]}{dt} = p'(y_t)\frac{dy_t}{dt} + \frac{C}{x_t^2}y_t\frac{dx_t}{dt}$$

We can use the first two lines of equation J.2 to eliminate $\frac{dR_t}{dt}$ and $\frac{dx_t}{dt}$, to write

$$R_t\left(r - \frac{dF(x_t)}{dx_t}\right) - \frac{C}{x_t^2}y_t = p'(y_t)\frac{dy_t}{dt} + \frac{C}{x_t^2}y_t\left(F(x) - y\right).$$

We can now use the third line of equation J.2 to eliminate R_t, to write

$$\left(p(y_t) - \frac{C}{x_t}\right)\left(r - \frac{dF(x_t)}{dx_t}\right) - \frac{C}{x_t^2}y_t = p'(y_t)\frac{dy_t}{dt} + \frac{C}{x_t^2}y_t\left(F(x) - y\right).$$

Solving this equation for $\frac{dy_t}{dt}$ gives

$$\frac{dy}{dt} = \frac{\left(p(y_t) - \frac{C}{x_t}\right)\left(r - \frac{dF(x_t)}{dx_t}\right) - \frac{C}{x_t^2}y_t - \frac{C}{x_t^2}y_t\left(F(x) - y\right)}{p'(y_t)} = H(x_t, y_t)$$

The middle expression is a function of only x and y, and the model parameters. We define this expression as the function $H(x_t, y_t)$.

This function looks complicated, but for the functional forms and the parameter values in our example, it simplifies to

$$\frac{dy}{dt} = 0.000\,02\frac{-500.0yx + 80.0yx^2 - 28.0x^2 + 195.0x + 750.0}{x} = H(x_t, y_t).$$

Chapter 16.3.2 uses this function to construct the y isocline, the set of points where $\frac{dy}{dt} = 0$. This isocline is given by

$$\frac{dy}{dt} = 0 \Rightarrow y = -\frac{1}{10x\,(0.08x - 0.5)}\left(-0.28x^2 + 1.95x + 7.5\right).$$

J.3 Finding the full solution

In order to find the solution to the optimization problem, needed to construct the dotted curve in Figure 16.3, we take the ratio of the differential equation for y and the differential equation for x, to obtain a new differential equation, showing how y changes with changes in x:

$$\frac{\frac{dy}{dt}}{\frac{dx}{dt}} = \frac{dy}{dx} = \frac{H\,(x_t y_t)}{F\,(x) - y}.$$

The solution to this equation is a function giving the optimal harvest as a function of the stock, the optimal "harvest rule". Denote this function as $y = Y\,(x)$. Figure 16.3 shows the graph of $Y\,(x)$, the dotted curve. Solving the differential equation to obtain the optimal harvest rule, $Y\,(x)$, requires a "boundary condition", giving the value of y at some value of x. Our boundary condition is given by the steady state, denoted (x_∞, y_∞). We calculate the steady state by finding the intersection of the x and the y isoclines. Our boundary condition is $y\,(x_\infty) = y_\infty$.

Some numerical algorithms encounter a problem in solving the differential equation, because both the numerator and denominator of $\frac{dy}{dx}$ vanish at the steady state, making the ratio an indeterminate form. This problem is easily resolved, but involves methods beyond the scope of this book. We can linearize our original non-linear system and use the eigenvector associated with the stable eigenvalue to replace the boundary condition (the steady state) with a point on the "stable" eigenvector. We have to (numerically) solve the resulting initial value problem twice, once beginning with a point slightly below the steady state, and then beginning with a point slightly above the steady state. Figure 16.3 shows the first of these two parts of the solution.

The dotted curve in Figure 16.3 shows the graph of the optimal harvest level, as a function of the stock, x. Harvest is positive only when the stock is above 3. As the stock increases over time, the harvest rises. The harvest is nearly constant, once the stock reaches about 20 or 25. The stock continues to grow to its steady state, and the harvest changes very little.

Appendix K

The common property water game

We define an individual farmer's benefit of consuming $\frac{y}{n}$ units of water as $v\left(\frac{y}{n}\right)$; when each of n farmers consumes $\frac{y}{n}$ units, the total benefit of consumption is $nv\left(\frac{y}{n}\right) \equiv V(y)$. With this definition, the net benefit to Farmer i is $v\left(y_t^i\right) - (c_0 - cx_t)\, y_t^i$, and the aggregate benefit, when each farmer consumes an equal share $(y^i = \frac{y}{n})$ equals $V(y_t) - (c_0 - cx_t)\, y_t$. Replacing $V(y)$ with $nv\left(\frac{y}{n}\right)$ does not alter the social planner's problem, or the Euler equation for that problem, but it provides the notation needed to think about the game when each farmer individually chooses her own extraction. In a symmetric equilibrium, each farmer has the same level of consumption in a period: $y_t^i = \frac{y_t}{n}$. Using the chain rule and $V(y) \equiv nv\left(\frac{y}{n}\right)$, we have $V'(y) = nv'\left(\frac{y}{n}\right)\frac{1}{n} = v'\left(\frac{y}{n}\right)$. When Farmer i's benefit of extraction is $v\left(y_t^i\right) - (c_0 - cx_t)\, y_t^i$, her rent in a symmetric equilibrium is

$$R_t^i\left(\frac{y_t}{n}, x_t\right) = v'\left(\frac{y_t}{n}\right) - (c_0 - cx_t) = V'(y) - (c_0 - cx_t) = R(y_t, x_t)\,. \quad \text{(K.1)}$$

These equalities state that for a *given level of extraction*, y_t, and a given stock, x_t, the individual farmer's rent and the social planner's rent are the same. Of course, the equilibrium level of extraction differs in a common property game and under the social planner.

Consider a noncooperative Nash equilibrium in which farmer i extracts y_t^i units of water at t, and takes as given the *aggregate* extraction policy (a function of the stock, x) of all *other* farmers.[1] We denote that aggregate

[1] There are a number of different types of Nash equilibria in dynamic games of this sort.

extraction policy as $y^*(x_t)$. Farmer i faces the equation of motion

$$x_{t+t} - x_t = F^*(x_t) - y_t^i, \text{ with } F^*(x_t) \equiv F(x_t) - y^*(x_t). \tag{K.2}$$

and her Euler equation is (cf. equation 17.6)

$$R_t^i = \rho \left(R_{t+1}^i \left(1 + \frac{dF^*(x_{t+1})}{dx_{t+1}} \right) + c y_{t+1}^i \right). \tag{K.3}$$

As in our two-period example, we want to compare the optimality conditions under the planner (equation 17.6) and in the game (equation K.3), without actually solving for the two equilibria. The left sides of these two equations are the same (by virtue of Equation K.1), but their right sides differ (just as is the case with the two first order conditions in our two-period example). The right side of equation 17.6 contains cy_{t+1}, accounting for the higher *aggregate* costs in period $t+1$ due to the lower stock. In contrast, the right side of equation K.3 contains cy_{t+1}^i, accounting for the higher cost only to Farmer i due to the lower stock. When the planner decides whether to extract an extra unit, she takes into account the higher aggregate future cost; the individual farmer only takes into account her own future higher cost. The higher cost that other farmers face is the "cost externality" discussed in the text.

The "scarcity externality" arises from the fact that the right side of the equation 17.6 contains the term $\frac{dF(x_{t+1})}{dx_{t+1}}$, whereas the right side of equation K.3 contains $\frac{dF^*(x_{t+1})}{dx_{t+1}}$. If Farmer i (irrationally) believes that the other farmers would not condition their future extraction decisions on the future water stock, then $y^{*\prime}(x) = 0$ and these two terms are identical. In that case, the Euler equation does not reflect a scarcity externality.[2] However, a reasonable conjecture for equilibrium is that

$$\frac{dy^*(x_t)}{dx} > 0. \tag{K.4}$$

This inequality states that a higher stock of water leads to higher extraction by the other agents. The assumption is reasonable, because the higher is

We consider a *feedback* (also known as Markov perfect) equilibrium, in which each agent thinks that all of the other agents will base their decisions on the "payoff relevant" state variable. Here, the payoff relevant state variable is the stock of water.

[2] Our analysis ignores the transversality condition. When the resource is eventually exhausted, the scarcity externality also shows up in this condition.

the stock of water, the lower are extraction costs, and the less scarce is the water. Both of these considerations tend to encourage higher extraction.

Inequality K.4 means that actions are "dynamic strategic substitutes", in the following sense: If agent i extracts an extra unit of water at time t, the stock in the next period will be lower than it otherwise would have been, causing other farmers' extraction decisions to be lower than they otherwise would have been. That is, higher extraction by farmer i at a point in time causes other farmers to reduce their future extraction.

If equation K.4 holds, then

$$\frac{dF^*(x_{t+1})}{dx_{t+1}} < \frac{dF(x_{t+1})}{dx_{t+1}}.$$

This inequality lowers the reduction in extraction that Farmer i needs to make at time $t+1$, following an increase in her extraction at t (in order to return to the candidate trajectory). By leaving her neighbors with a lower stock, Farmer i induces them to lower their future extraction, benefitting Farmer i. The neighbors' future response to lower stocks encourages Farmer i to increase her current extraction.

Appendix L

Sustainability

We first confirm the Hartwick Rule and then examine the feasibility of sustainability, when society follows the Hartwickt rule.

L.1 Confirming the Hartwick Rule

Here we show that the Hotelling Rule + Hartwick Rule implies constant consumption ($\frac{dC}{dt} = 0$). Reordering the argument shows that constant consumption + the Hotelling Rule implies the Hartwick Rule.

The national income accounting identity states that total income (Y) must equal total expenditures. Expenditure is the sum of investment ($I = \frac{dK}{dt}$) and extraction costs (cE) and consumption (C):

$$\text{Income accounting identity:} \quad Y = \frac{dK}{dt} + cE + C \tag{L.1}$$

We rearrange this identity to write.

$$Y - (I + cE) = C$$

Using the Hartwick Rule, $I = (p - c)\,E$, we have

$$Y - ((p - c)\,E - cE) = Y - pE = C. \tag{L.2}$$

Differentiating both sides with respect to time gives

$$\frac{dY}{dt} - \frac{d\,(pE)}{dt} = \frac{dC}{dt}.$$

We use the differential for $Y(K,E)$ to write the left side of this equation as

$$F_K I + F_E \frac{dE}{dt} - \frac{d(pE)}{dt} = r(p-c)E + p\frac{dE}{dt} - \frac{d(pE)}{dt}.$$

The equality uses the Hartwick Rule and equation 18.1, the fact that the value of marginal product equals factor price. Using the Hotelling Rule, we write the right side of the last expression as

$$\frac{dp}{dt}E + p\frac{dE}{dt} - \frac{d(pE)}{dt} = 0.$$

The equality follows from the product rule for differentiation. Thus, we have shown that the Hotelling Rule plus the Hartwick rule implies that consumption is constant over time.

L.2 Feasibility of constant consumption

Here we assume that technology is Cobb Douglas, $F(K,E) = K^{1-\alpha}E^{\alpha}$, a stronger assumption than constant returns to scale. We show that sustainable consumption is feasible if and only if $\alpha < 0.5$. As a preliminary step, we establish that under the Hartwick Rule, consumption is constant if and only if output, Y, is also constant. To demonstrate this claim, use the equilibrium condition that the value of marginal product of E equals the price of E:

$$\frac{\partial K^{1-\alpha}E^{\alpha}}{\partial E} = p \Rightarrow \alpha K^{1-\alpha}E^{\alpha-1} = p \Rightarrow pE = \alpha K^{1-\alpha}E^{\alpha} \Rightarrow \frac{pE}{Y} = \alpha. \quad \text{(L.3)}$$

The last equality states that payments to the resource sector, pE, as a share of the value of output, $K^{1-\alpha}E^{\alpha}$, equals the constant α. Using the last parts of equation L.2 and L.3, we have

$$Y = C + pE = C + \frac{pE}{Y}Y = C + \alpha Y \Rightarrow (1-\alpha)Y = C$$

The last equality implies that output (= income) is constant if and only if consumption is constant.

In order to determine whether a constant consumption path (i.e. a constant output path) is feasible, we solve $Y = K^{1-\alpha}E^{\alpha}$ for E to obtain

$E = Y^{\frac{1}{\alpha}} K^{\frac{\alpha-1}{\alpha}}$. For $\alpha \neq 0.5$, the integral of this function from an initial capital stock k to a larger stock z gives

$$R(z,k) \equiv Y^{\frac{1}{\alpha}} \int_k^z \left(K^{1-\frac{1}{\alpha}} \right) dK = \frac{\alpha}{2\alpha - 1} Y^{\frac{1}{\alpha}} \left(z^{2-\frac{1}{\alpha}} - k^{2-\frac{1}{\alpha}} \right).$$

For $\alpha = 0.5$, this integral is $R(z,k) = Y^2 \int_k^z (K^{-1}) \, dK = Y^2 (\ln z - \ln k)$. The function $R(z,k)$ equals cumulative extraction needed to produce a constant output Y as K varies from the initial level k to some larger level z. As noted in the text, capital becomes infinitely large along the sustainable trajectory, so (for $\alpha \neq 0.5$) a sustainable trajectory requires an initial resource stock of

$$\lim_{z\to\infty} R(z,k) = \lim_{z\to\infty} \frac{\alpha}{2\alpha-1} Y^{\frac{1}{\alpha}} \left(z^{2-\frac{1}{\alpha}} - k^{2-\frac{1}{\alpha}} \right) = \begin{cases} \infty \text{ if } \alpha > 0.5 \\ \frac{\alpha}{1-2\alpha} Y^{\frac{1}{\alpha}} k^{2-\frac{1}{\alpha}} \text{ if } \alpha < 0.5 \end{cases}.$$

For $\alpha = 0.5$, the initial resource stock needed in order to maintain constant output is $\lim_{z\to\infty} Y^2 (\ln z - \ln k) = \infty$. Thus, if $\alpha \geq 0.5$, it is not feasible to maintain *any* positive constant level of output, simply because such a path would require an infinite resource stock. If $\alpha < 0.5$, and the initial resource stock is x and the initial capital stock is k, it is feasible to maintain the constant level of output y that solves

$$x = \frac{\alpha}{1-2\alpha} Y^{\frac{1}{\alpha}} k^{2-\frac{1}{\alpha}} \Rightarrow Y = \left(\frac{(1-2\alpha)\, x}{\alpha} \right)^{\alpha} k^{1-2\alpha}.$$

For $Y = 1$, $k = 0.7$ and $\alpha = 0.4$ (as in Figure 18.1), $x = 2.4$. If $\alpha = 0.4$, the initial resource stock is 2.4 and the initial capital stock is 0.7, the constant output path $Y = 1$, and the corresponding consumption path $(1 - 0.4)\,1 = 0.6$ are sustainable.

Appendix M

Discounting

We derive the Ramsey formula for the consumption discount rate, and then discuss a numerical example that shows the effects, on willingness to pay to avoid future damages, of excessive optimism or pessimism.

M.1 Derivation of equation 19.1

We want to know how many units of consumption people today (time 0) are willing to sacrifice to increase time t consumption by 1 unit (one dollar or one billion dollars, depending on choice of units). Suppose, absent the policy, that society has c_0 units of consumption today for the present value utility $e^{-\rho\times 0}u(c_0) = u(c_0)$, and society has c_t units of consumption at time $t > 0$, with present value utility $e^{-\rho t}u(c_t)$. The utility discount factor $e^{-\rho t}$ converts the time t utility into its present value (at time 0, today) equivalent. If society gives up \$$x$ today, the utility cost is $u'(c_0)\,x$, the marginal value of a unit of consumption, times the number of units that society gives up today. The present value of the increased utility due to the extra dollar at time t is $e^{-\rho t}u'(c_t)$. Equating the marginal cost to the marginal gain gives

$$x(t) = \frac{e^{-\rho t}u'(c_t)}{u'(c_0)}.$$

This value of $x(t)$ equals the number of units of consumption society is willing to give up today, in exchange for one more unit of consumption at time t; $x(t)$ therefore is the *consumption discount factor*, giving the present value today of a future unit of consumption.

A rate of change (with respect to time) of a variable equals the derivative of the variable with respect to time, divided by the variable. Because $x(t)$ is the consumption discount factor, the absolute value of its rate of change is the consumption discount rate, which we denote as $r(t)$. Taking the derivative gives

$$\begin{aligned} r(t) &= -\frac{\frac{dx(t)}{dt}}{x(t)} = -\frac{\frac{d\left(e^{-\rho t}u'(c_t)\right)}{dt}}{e^{-\rho t}u'(c_t)} = \frac{\rho e^{-\rho t}u'(c_t) - e^{-\rho t}u''(c_t)\frac{dc}{dt}}{e^{-\rho t}u'(c_t)} \\ &= \frac{\rho e^{-\rho t}u'(c_t) - e^{-\rho t}u''(c_t)\,c\frac{\frac{dc}{dt}}{c}}{e^{-\rho t}u'(c_t)} = \rho - \frac{u''(c_t)\,c}{u'(c_t)}\frac{\frac{dc}{dt}}{c} = \rho + \eta_t g_t. \end{aligned}$$

The last equality uses the definitions in the second line of equation 19.1

M.2 Optimism versus pessimism about growth

Growth is $g(t) = \frac{.02}{1+\gamma t}$, $\gamma \geq 0$. The parameter γ determines growth's speed of decrease. For $\gamma = 0$, growth is constant at 2% per year; as $\gamma \to \infty$, growth falls almost immediately from 2% to 0%. We also use the intermediate value $\gamma = 0.0133$, for which annual growth falls to 1% after 75 years, and then gradually falls to 0. This example is broadly consistent with some complex policy-driven models, for which the current growth rate is 1.5% – 2%, and is expected to decline over time. Our example assumes that the true value is $\gamma = 0.0133$; $\gamma = 0$ implies "false optimism" and $\gamma = \infty$ implies "false pessimism" about growth.

If the CDR is constant, at r, then the consumption discount factor is e^{-rt}. If, instead, the CDR is a function of time, $r(t)$, then the consumption discount factor for a future time, t, is $e^{-R(t)t}$, with $R(t)$ equal to the average discount rate from today (time 0) and time t:

$$R(t) = \frac{\int_0^t r(\tau)\,d\tau}{t}.$$

The consumption discount factor, used to evaluate an exchange between the present and t periods in the future, depends on the consumption discount rates at all intervening periods. Using $\rho = 0.01$, $\eta = 2$, and $g(t) = \frac{.02}{1+\gamma t}$ with $\gamma = 0.0133$, the Ramsey formula implies $r(t) = 0.01 + 2 \times \frac{.02}{1+0.0133t}$; it falls over time from 5% to 1%, reaching the intermediate 3% level after 75 years.

Under growth-optimism ($\gamma = 0$), $r(t) = 0.05$ and under growth pessimism ($\gamma = \infty$), $r(t) = 0.01$. The short run growth predictions are similar under $\gamma = 0$ and $\gamma = .0133$ and very different for $\gamma = \infty$. In contrast, the long run growth predictions are similar under $\gamma = 0.0133$ and $\gamma = \infty$, and very different under $\gamma = 0$.

The three scenarios with $\gamma = 0$ (growth is constant at 2%), $\gamma = 0.0133$ (described above), and $\gamma = \infty$ (future growth is zero) illustrate the policy importance of assumptions about growth over long stretches of the future. For each of these scenarios, we ask "What is the maximum risk premium (measured in dollars) that society would pay, in perpetuity, in order to avoid a \$100 perpetual loss in consumption beginning T years in the future?"[1] Denote this Willingness to Pay as $WTP(\gamma, T)$, a function of T and γ.

To show the policy relevance of assumptions about future growth, we consider the ratios of $WTP(\gamma, T)$ for different values of γ and T. Denote

$$\text{Ratio}(1,T) = \frac{WTP(\gamma = 0.0133, T)}{WTP(\gamma = 0, T)}; \; \text{Ratio}(2,T) = \frac{WTP(\gamma = \infty, T)}{WTP(\gamma = 0.0133, T)}$$

For example, if Ratio(1,T) = 10, then the planner is willing to spend 10 times the amount to avoid the event when growth falls ($\gamma = 0.0133$) compared to when growth is constant at 2% ($\gamma = 0$). Because we assume that $\gamma = 0.0133$ describes actual growth, Ratio(1,T) equals the magnitude of the error if we are too optimistic about growth, and Ratio(2,T) equals the error if we are too pessimistic. The "error" is the understatement or overstatement of WTP, relative to the correct WTP when we know $\gamma = 0.0133$.

Figures M.1 and M.2 show graphs of these two ratios as functions of the event time, T. The first figure graphs these two ratios as T varies from 0 to 120 years, and the second figure shows the ratios as T varies from 120 to 220 years. By using two figures, we can see how the scale of the comparison depends on the event time, T. For example:

- If the event time is $T = 50$, Ratio(1) = 2.2 and Ratio(2) = 3.3. In this case, the error (in calculating the correct Willingness to Pay) arising from to being too pessimistic is $\frac{3.3}{2.2}100 = 150\%$ of the error arising from being too optimistic.

[1] Chapter 19.1.1 addresses a similar question, but here we measure the trade-off in dollars instead of utility, and we take into account the possibility of growth.

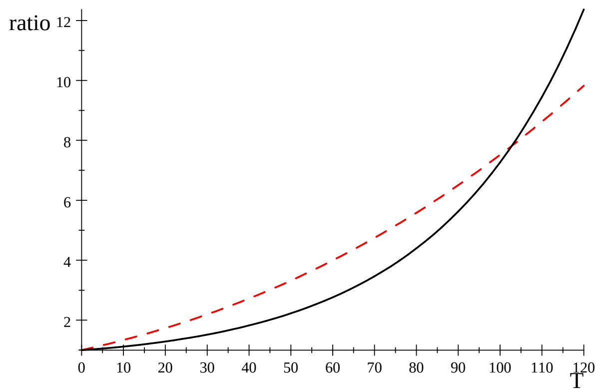

Figure M.1: Solid graph: Ratio $(1,T)$; Dashed graph: Ratio $(2,T)$.

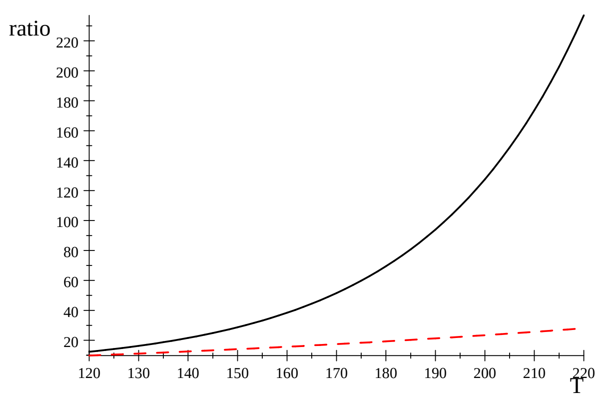

Figure M.2: Solid graph: Ratio $(1,T)$; Dashed graph: Ratio $(2,T)$.

- If the event time is $T = 200$, Ratio(1) $= 127.5$ and Ratio(2) $= 23.4$. In this case, the error (in calculating the true Willingness to pay) arising from being too optimistic is $\frac{127.5}{23.4}100 = 544\%$ of the error arising from being too pessimistic.

This example illustrates that the cost-benefit analysis of a public investment with a payoff in the near future, e.g. the next century, depends largely on near-term growth. In contrast, the cost benefit analysis of a public investment with a payoff in the distant future is much more sensitive to growth rates over long spans of future time. We probably know much less about growth in the distant compared to the near future. Overestimates of future growth lead to too low an estimate of willingness to pay to avoid future damages. Underestimates of future growth lead to too high an estimate of willingness to pay.

Afterword

The back matter often includes one or more of an index, an afterword, acknowledgements, a bibliography, a colophon, or any other similar item. In the back matter, chapters do not produce a chapter number, but they are entered in the table of contents. If you are not using anything in the back matter, you can delete the back matter TeX field and everything that follows it.

Bibliography

Abbott, J., and J. Wilen (2009): "Regulating of fisheries bycatch with common-pool output quoatas," *Journal of Environmental Economics and Management*, 57, 195–204.

——— (2011): "Dissecting the tragedy: a spatial model of behavior in the commons," *Journal of Environmental Economics and Management*, 62, 386–401.

Administration, U. E. I. (2014): "Annual Energy Outlook 2014," DOE/IEA-0383(2014).

Aker, J. (2010): "Information from markets near and far: Mobile phones and agricultural markets in Niger," *American Economic Journal: Applied Economics*, 2, 46–59.

Aldy, J. (2013): "Eliminating Fossil Fuel Subsidies," The Hamilton Project, Brookings.

Alix-Garcia, J., A. de Janvry, E. Sadoulet, and J. Torres (2009): "Payment for eco-system services program," in *Payments for Environmental Services in Agricultural Landscapes*, ed. by L. Lipper, T. Sakuyama, R. Stringer, and D. Zilberman, pp. 163 – 188. FAO and Springer.

Aranson, R. (2012): "Property rights in fisheries: how much can Individual Transferable Quotas accomplish?," *Review of Environmental Economics and Policy*, 10.1093/reep/res011.

Arnason, R., K. Kelleher, and R. Willmann (2009): "The sunken billioms: the economic justification for fisheries reform," The World Bank, Agriculture and Rural Development.

Arrow, K., and 13 others (2012): "How should benefits and costs be discounted in an intergenerational context?," Resources for the Future Discussion Paper 12 - 53.

Assessment, I. C. (2012): "Global Water Security," ICA 2012-08, 2 February 2012.

Auerbach, A., and J. Hines (2002): "Taxation and Economic Efficienc, vol. 1y," in *Handbook of Public Economics*, ed. by A. Auerbach, and M. Feldstein, pp. 1347–1421. North Holland, Amsterdam.

Baerenklau, K., K. Schwabe, and A. Dinar (2015): "The residential water demand effect of increasing block rates water budgets," *Land Economics*, 90, 683 – 699.

Barbier, E. (2011): *Capitalizing on Nature: Ecosystems as Natural Assets.* Cambridge University Press.

Barrett, S. (2003): *Environment and Statecraft.* Oxford University Press.

Barro, R. (2006): "Rare disasters and asset merkets in the twentieth century," *Quarterly Journal of Economics*, 121, 823 – 866.

Berck, P., and G. Helfand (2010): *The Economics of the Environment.* Pearson.

Berck, P., and J. Perloff (1984): "An open access fishery with rational expectations," *Econometrica*, 52, 489 – 506.

Berck, P., and M. Roberts (1996): "Natural-resource prices: Will they ever turn up?," *Journal of Environmental Economics and Management*, 31, 65 – 78.

Boadway, R., and M. Keen (2009): "Theoretical perspectives on resource tax design," Queens Economics Department Working Paper, No 1206.

Bohn, H., and R. Deacon (2000): "Ownership risk, investment, and the use of natural resources," *American Economic Review*, 90, 526–49.

Bolster, J. (2015): "Where have all the cod gone?," Opinion Pages, New York Times January 2.

BOVENBERG, A., AND F. VAN DER PLOEG (1994): "Environmental policy, public finance, and the labour market in a second-best world," *Journal of Public Economics*, 55, 349 – 390.

BRAXTON, J. (2009): "The Ogallala Aquifer: Saving a vital U.S. water source," Scientific American, March 2009.

BROMLEY, D. (2015): "Correcting the whimsies of U.S. fisheries policies," *Choices*, 30, 1–7.

BRYCE, R. (2015): "The hidden corn ethanol tax – How much does the Renewable Fuel Standard costs motorists?," Manhattan Institute Issues Brief 32.

BULTE, E., AND E. BARBIER (2005): "Trade and Renewable Resources in a Second Best World: An Overview," *Environmental and Resource Economics*, 30, 423–463.

CD DII MARIA, I. L., AND E. VAN DER WERF (2013): "Goiong full Circle: Demand-Side Constraints to the Green Paradox," CESifo Working per No. 4152.

CERDA, R. (2007): "Market power and primary commodity prices: the case of copper," *Applied Economics Letters*, 14, 775–778.

CHERMAK, J. M., AND R. PARTICK (1995): "A well-based cost function and the economics of exhaustible resources: The case of natural gas," *Journal of Environmental Economics and Management*, 28, 174–189.

——— (2001): "A microeconometric test of the theory of exhaustible resources," *Journal of Environmental Economics and Management*, 42, 82 – 103.

——— (2002): "Comparing tests ofthe theory of exhaustible resources," *Resource and Energy Economics*, 24, 301–325.

CHICHILNISKY, G. (1994): "North-South Trade and the Global Environment," *American Economic Review*, 84(4), 851–874.

CHONG, H., AND D. SUNDING (2006): "Water Markets and Trading," *Annual Review of Environment and Resources*, 31, 239–264.

Chu, C. (2009): "Thirty years later: the global growth of ITQs and their influences ion stock status in marine fisheries," *Fish and Fisheries*, 10, 217–230.

Clark, C. (1996): *Mathematical Bioeconomics.* Wiley Interscience, New York.

Coady, D., I. Parry, L. Sears, and B. Shang (2011): "How large are global energy subsidies?," IMF Working Paper WP/15/105.

Conover, C. (2010): "Congress should account for the excess burden of taxation," Cato Institute, Policy Analysis, October 13, 2010.

Conrad, J. (2010): *Resource Economics, 2nd ed.* Cambridge University Press, Cambridge University Press.

Copeland, B. R., and M. S. Taylor (2009): "Trade, Tragedy, and the Commons," *American Economic Review*, 99, 725–749.

Costello, C. (2012): "Introduction to the symposium on rights-based fisheries management," *Review of Environmental Economics and Policy*, 10.1093/reep/res011.

Costello, C., S. Gaines, and J. Lynham (2008): "Can catch shares prevent fisheries collapse," *Science*, 321, 1678–1681.

Dasgupta, P. (2001): *Human Well-Being and the Natural Environment.* Oxford University Press.

Dasgupta, P., and G. Heal (1974): "The optimal depletion of exhaustible resources," *The Review of Economic Studies*, 41, 3–28.

——— (1979): *Economic Thoery and Exhaustible Resources.* Cambridge University Press, Cambridge.

Daubanes, J., and S. A. de Sa (2014): "Taxing the rent of non-renewable resource sectors: a theoretical note," OECD Economics Department Working Papers, No 1149.

Deacon, R. (2012): "Fishery management by harvester coops," *Review of Environmental Economics and Policy*, 10.1093/reep/res011.

DEACON, R., D. PARKER, AND C. COSTELLO (2013): "Reforming fisheries: lessons from a self-selected coopeartive," *Journal of Law and Economics*, 56, 83–125.

DIETZ, T., E. OSTROM, AND P. C. STERN (2003): "The Struggle to Govern the Commons," *Science*, 302(5652), 1907–1912.

DIEWERT, W. E., D. A. LAWRENCE, AND F. THOMPSON (1998): "The marginal cost of taxation and regulation," in *Handbook of Public Finance*, ed. by F. Thomson, and M. T. Green. Marcel Decker, New York.

DYCK, A., AND U. R. SUMAILA (2010): "Economic impact of ocean fish populations in the global fishery," *Journal of Bioeconomics*, 12, 227–43.

EKELAND, I., AND A. LAZRAK (2010): "The golden rule when preferences are time inconsistent," *Mathematical and Financial Economics*, 4(1), 29–55.

ELLERMAN, A., AND J. MONTERO (1998): "The declining trend in sulfur dioxide emissions: implications for allowance prices," *Journal of Environmental Economics and Policy*, 36, 26–45.

ELLIS, G., AND R. HALVORSEN (2002): "Estimation of market power in a nonrenewable resource industry," *The Journal of Political Economy*, 110, 883–899.

EPSTEIN, L., AND S. ZINNL (1991): "Substitutition, risk aversion, and the temporal behavior of consumption and asset returns," *Journal of Political Economyl*, 99, 263 – 286.

FAMA, E., AND K. FRENCH (2004): "The Capital Asset Pricing Model: Theory and evidence," *Journal of Economic Perspectives*, 18, 25–46.

FARZIN, Y. (1984): "The effect of the discount rate on depeletion of exhaustible resources," *Journal of Political Economy*, 92, 841–51.

——— (1995): "Technological change and the dynamics of resource scarcity measures," *Journal of Environmental Economics and Management*, 29, 105–120.

FELDSTEIN, M. (1977): "The surprising incidence of a tax on pure rent: a new answer to an old question," *Journal of Political Economy*, 85, 349–360.

Fenichel, E. P., and J. K. Abbott (2014): "Natural Capital: From Metaphor to Measurement," *Journal of the Association of Environmental and Resource Economics*, 1, 1–27.

Fenske, J. (2012): "Imachi Nkwu: trade and the commons," CSAE Working Paper WPS/2012-19.

Fisher, A. (1981): *Resource and Environmental Economics*. Cambridge University Press.

Fowlie, M. (2009): "Incomplete Environmental Regulation, Imperfect Competition, and Emissions Leakage," *American Economic Journal: Economic Policy*, 1, 72–112.

Fowlie, M., S. Holland, and E. Mansur (2012): "What do emissions markets deliver and to whome? Eveidence from Southern California's NOx trading program," *American Economic Review*, 102, 965 –993.

Fowlie, M., and J. Perloff (2013): "Distributing pollution rights in Cap-and-Trade programs: are outcomes independent of allocation?," *Review of Economics and Statistics*, 95, 1640–1652.

Gentry, W. (2007): "A review of the evidence of the incidence of the Corporate Income Tax," US Department of Treasury, OTA Paper 101.

Gerlagh, R., and M. Liski (2012): "Carbon prices for the next thousand years," Unpublished Working Paper.

Gerlgh, R. (2011): "Too much oil," *CESifo Economic Studies*, pp. 79 – 102.

Glaeser, E., and J. Kohlhase (2004): "Cities, regions, and the decline of transport costs," *Papers in Regional Science*, 83, 197–228, DOI: 10.1007/s10110-003-0183-x.

Gollier, C. (2014): "Evaluation of long-dated investments under uncertain growth trend, volatility, and catastrophes," Toulouse School of Economics Working Paper.

Gordon, H. S. (1954): "The Economic Theory of a Common-Property Resource: The Fishery," *Journal of Political Economy*, 62, 124–142.

Gordon, R. (2015): *The Rise and Fall of American Growth.* Princeton University Press.

Goulder, L. (1995): "Environmental taxation and the double dividend," *International Tax and Public Finance*, pp. 157 – 183.

grafton, R. Q., T. Kompas, and R. Hilborn (2007): "Economics of overexploitation revisited," *Science*, pp. 1601 – 1604.

Guadet, G. (2007): "Natural resource economics under the Rule of Hotelling," Presidential address, Canadian Economics Association.

Halvorsen, R., and T. Smith (1991): "A test of the theory of exhaustible resources," *Quarterly Journal of Economics*, 106, 123–140.

Hamilton, J. (2011): "Nonlinearities and the macroeconomic effects of oil prices," *Macroeconomic Dynamics*, 15, 364–378.

Hardin, G. (1968): "The Tragedy of the Commons," *Science*, 162(3859), 1243–1248.

Hartwick, J. (1977): "Intergenerational equity and the investing of rents from exhaustible resources," *American Economic Review*, 67, 972–974.

——— (2011): "Green National Income and Green National Product," *Annual Review of Resource Economics*, 3, 21–35.

Hartwick, J., and N. Olewiler (1986): *The Economics of Natural Resource Use.* Harper and Row, New York.

Heal, G., and M. Barrow (1980): "The relationship between interest rates and metal-price movements," *Review of Economic Studies*, 47, 161–182.

Hinterman, B. (2015): "Market poer in emission permit markets: Theory and evidence from the EU ETS," *Environmental and Resource Economics*, p. xx, DOI 10.1007/s10640-015-9939-4.

Hnyilicza, E., and R. Pindyck (1976): "Pricing policies for a two-part exhaustible resource cartel: the Case of OPEC," *European Economic Review*, 8, 139–154.

Hoel, M. (2008): "Bush meets Hotelling: the effects of improved renewable energy technology on greenhouse gas emissions," University of Oslo Working Paper.

——— (2012): "The supply side of CO2 with country heterogeneity," *Scandanavian Journal of Economics*, 113, 846–865.

Holland, S. (2012): "Emissions taxes versus intensity standards: Second-best environmental policies with incomplete regulation," *Journal of Environmetnal Economics and Management*, 63, 375–87.

Holland, S., J. Hughes, C. Knittel, and N. Parker (2015): "Some inconvenient truths about climate change policy: the distributional impacts of tranportation policies," *Review of Economics and Statistics*, p. xxx.

Homans, F. R., and J. Wilen (1997): "A model of regulated open access resource use," *Journal of Environmental Economics and Management*, 32, 1–21.

——— (2005): "Markets and rent dissipation in regulated open access fisheries," *Journal of Environmental Economics and Management*, 49, 381–404.

Hotelling, H. (1931): "The economics of exhaustible resources," *Journal of Political Economy*, 39, 137–175.

Howitt, R. (1994): "Empirical analysis of water market institutions," *Resource and Energy Economics*, 16, 357–371.

i, A. A. (2015): "The Ganges water crisis," New York Times, June 17.

IEA, OPEC, OECD and World Bank (2011): "Joint report by IEA, OPEC, OECD and World Bank on fossil-fuel and other subsidies," .

Interagency Working Group (2013): "Technical update of the Social Cost of Carbon," .

International Scientific Committee for Tuna (2011): "Report of the Albacore Working Group Stock Assessment Workshop," .

IVERSON, T. (2015): "Optimal carbon taxes with non-constant time preference," http://mpra.ub.uni-muenchen.de/43264/.

JOHNSON, R., AND G. LIBCAP (1982): "Contracting problems and regulation: the case of the fisheryl," *American Economic Review*, 72, 1005 – 1022.

JUDD, K. (1985): "Redistributive taxation in a simple perfect foresight model," *Journal of Public Economics*, 28, 59–83.

KAFFINE, D. (2003): "Quality and the Commons: the surf gangs of California," *Journal of Law and Economics*, 52, 727–743.

KAMIEN, M., AND N. SCHWARTZ (1991): *Dynamic Optimization, 2nd edition*. North Holland.

KARP, L. (2005): "Global warming and hyperbolic discounting," *Journal of Public Economocs*, 89, 261–282.

——— (2016): "Discounting utility and the evaluation of climate policy," in *The Economics of the global environment - Catastrophic risks in theory and practice*, ed. by G. Chichilnisky, and A. Rezai. Springerr.

KARP, L., AND I. H. LEE (2003): "Time Consistent Policies," *Journal of Economic Theory*, 112(2), 353–364.

KARP, L., AND A. REZAI (2015): "Trade and resource sustainability with asset markets," Working paper.

KARP, L., AND M. STEVENSON (2012): "Green Industrial Policy : Trade and Theory," World Bank Working Paper, https://openknowledge.worldbank.org/handle/10986/12081.

KASSLER, D. (2015): "California Supreme Court wont budge on water rates," Sacramento Bee July 23.

KILIAN, L. (2009): "Not all oil price shocks are alike: Disentangling demand and supply shocks in the crude oil market," *American Economic Review*, 99, 1053–1069.

KRAUSE, S., F. BOANO, M. CUTHBERT, J. FLECKENSTEIN, AND J. LEWANDOWSKI (2014): "Understanding process dynamics at aquifer-surface water interfaces: An introduction to the special section on new modeling approaches and novel experimental technologies," *Water Resources Research*, 50, 1847–1855.

KRAUTKRAEMER, J. (1998): "Nonrenewable resource scarcity," *Journal of Economic Literature*, 36, 2065–2107.

KRONENBERG, T. (2008): "Should we worry about failure of the Hotelling Rule?," *Journal of Economic Surveys*, 22, 774–793.

KURLANSKY, M. (1998): *Cod: a Biography of the Fish that Changed the World*. Penguin Books.

LAIBSON, D. (1997): "Golden eggs and hyperbolic discounting," *Quarterly Journal of Economics*, 62, 443–78.

LALL, R. (2004): "Soil Carbon Sequestration Impacts on Global Climate Change and Food Security," *Science*, 304, 1623–27.

LEE, J., J. A. LIST, AND M. STRAZICICH (2006): "Nonrenewable resource prices: deterministic or stochastic trends?," *Journal of Environmental Economics and Management*, 51, 354–370.

LEMOINE, D. (2015a): "The climate risk premium," Working paper, University of Arizona.

——— (2015b): "Green Expectations: current effects of anticipated carbon pricing," Working paper, University of Arizona.

LEONHARDT, D. (2012): "A Climate proposal beyond cap and trade," New York Times, October 12.

LIN, C., AND G. WAGNER (2007): "Steady state growth in a Hotelling Model of resource extraction," *Journal of Environmental Economics and Management*, 54, 68–83.

LIPSEY, R., AND K. LANCASTER (1956): "The General Theory of Second BestF," *Review of Economic Studies*, 24, 11–32.

LIVERNOIS, J. (1987): "Empirical evidence on the characteristics of extractive technologies: the case of oil," *Journal of Environmental Economics and Management*, 14, 72–86.

——— (2009): "On the emprical significance of the Hotelling rule," *Review of Environmental Economics and Policy*, 3, 21–41.

LIVERNOIS, J., H. THILLE, AND X. ZHANG (2006): "A test of the Hotelling Rule using old-growth timber," *Canadian Journal of Economics*, 39, 163–186.

LIVERNOIS, J., AND R. UHLER (1987): "Extraction costs and the economics of nonrenewable resources," *Journal of Political Economy*, 95, 1195–203.

LLAVADOR, H., J. ROEMER, AND J. SILVESTRE (2016): *Sustainability for a warming planet*. Harvard University Press.

LUND, D. (2009): "Rent taxation for nonrenewable resources," *Annual Review of Resource Economics*, 1, 10.1146/annurev.resource.050708.144216.

MALISCHEK, R., AND C. TODE (2015): "A test of the theory of nonrenewable resouces - controlling for exploration and market power," Institute of Energy Economics, University of Cologne.

MANGEL, M. (1985): *Decision and Control in Uncertain Resource Systems*. Academic Press.

MCAUSLAND, C. (2003): "Voting for Pollution Policy: The Importance of Income Inequality and Openness to Trade," *Journal of International Economics*, 61(2), 425–51.

——— (2008): "Trade, Politics and the Environment: Tailpipe vs.Smokestack," *Journal of Environmental Economics and Management*, 55, 52–71.

MILLER, M., AND C. UPTON (1985): "A test of the Hotelling valuation principle," *Journal of Political Economy*, 93, 1–25.

NAIR, C. (2014): "Tackling a drought in Iran," New Your Times November 4.

NATIONAL RESEARCH COUNCIL (2011): *Renewable Fuel Standard: Potential economic and environmental effects of U.S. Biofuel Policy*. National Academies Press.

NEWELL, R., J. SANCHIRICO, AND S. KERR (2005): "Fishing quota markets," *Journal of Environmental Economics and Management*, 49, 437–462.

OKUM, A. (1975): *Equality and Efficiency: The Big Tradeoff*. The Brookings Institution.

OSTROM, E. (1990): *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge University Press.

——— (2007): "A Diagnostic Approach for Going beyond Panaceas," *Proceedings of the National Academy of Sciences*, 104(39), 15181–15187.

PINDYCK, R. (1978): "The optimal exploration and production of nonrenewable resources," *Journal of Political Economy*, 86, 841–861.

——— (1980): "Uncertainty and exhaustible resource markets," *Journal of Political Economy*, 86, 841–61.

——— (1987): "On monopoly power in extractive resource markets," *Journal of Environmental Economics and Management*, 14, 128–142.

——— (1999): "The long-run evolution of energy prices," *Energy Journal*, 20, 1–27.

PRAKASH, A., AND M. POTOSKI (2007): "Collective Action through Voluntary Environmental Programs: A Club Theory Perspective," *Policy Studies Journal*, 35, 773– 792.

RAMSEY, F. (1928): "A Mathematical Theory of Savings," *Economic Journal*, 38, 543–59.

REUTERST (2014): "Shell says fossil fuel reserves won't be 'stranded' by climate regualtion," May 19.

SCHAEFER, M. (1954): "Some aspects of the dynamics of population important to the management of commercial marine fisheries," *Bulletin of the InterAmerican Tropical Tuna Commission*, 1, 26–56.

Schoengold, K., and D. Zilberman (2007): “The economics of water, irrigation, and development,” in *Handbook of Agricultural Economics*, ed. by R. Evenson, and P. Pingali, pp. 2933 – 2977. Elsevier.

Schwartzl, J. (2015): “Norway will divest from coal in push against climate change,” New York Times June 5.

Schwartzmayl, N. (2015): “Water pricing in two thirsty cities: in one, guzzlers pay more and use less,” New York Times May 6.

Sen, A. (1987): *On Ethics and Economics*. Blackwell.

Service, N. M. F. (2009): “Status of Fisheries of the United States: A Report to Congress,” .

Sharp, R., and U. Sumaila (2009): “Quantification of U.S. marine subsidies,” *North American Journal of Fisheries Management*, 29, 18 – 32.

Sijm, J., K. Neuhoff, and Y. Chen (2006): “CO2 cost pass through and windfall profits in the power sector,” *Climate Policy*, 6, 49–72.

Sinn, H. (2008): “Public Policies against Global Warming: A Supply Side Approach,” *International Tax and Public Finance*, 15, 360–394.

Sinn, H. W. (2012): *The Green Paradox*. Massachusetts Institute of Technology.

Slade, M. (1984): “Tax policy and the supply of exhaustible resources,” *Land Economics*, 60, 133 – 47.

Slade, M., and H. Thille (1997): “Hotelling confronts CAPM: a test of the theory of exhaustible resources,” *Canadian Journal of Economics*, 30, 685 – 708.

——— (2009): “Whither Hotelling: Tests of the Theory of Exhaustible Resources,” *Annual Review of Resource Economics*, 1, 239–259.

Smith, M. (2012): “The new fisheries econom: Incentives across many margins,” *Annual Review of Resource Economics*, 4, 379–402.

Solow, R. (1974a): “The economics of resources or the resources of economics,” *Review of Economic Studies*, pp. 29–46.

——— (1974b): "Intergenerational equity and exhaustible resources," *American Economic Review*, 64, 1–14.

Squires, D., and N. Vestergaard (2013): "Technical change and the commons," *The Review of Economics and Statistics*, 95, 1769–1787.

Stewarda, D., P. Brussa, X. Yangb, S. Staggenborgc, S. Welchc, and M. Apleyd (2013): "Tapping unsustainable groundwater stores for agricultural production in the High Plains Aquifer of Kansas, projections to 2110," *PNAS*, 110, www.pnas.org/cgi/doi/10.1073/pnas.1220351110.

Stollery, K. R. (2008): "Mineral depletion with cost as the extraction limit," *Journal of Environmental Economics and Management*, 10, 151–65.

Sumailia, U., W. Cheung, V. Lam, D. Pauly, and S. Herrick (2011): "Climate change impacts on the biophysics and economics of world fisheries," *Nature Climate Change*, doi:10.1038/nclimate1301.

Sumailia, U., V. Lam, F. Le Manach, W. Wartz, and D. Pauly (2013): "Global Fisheries Subsidies," Report to European Parliament.

Tietenberg, T. (2006): *Environmental and Natural resource Economics, 7th ed.* Pearson, Boston.

Timilsinia, G., L. Kurdgelashvili, and P. Narbel (2011): "A review of solar energy," The wold Bank Policy Research working Papaer 5845.

Traegerl, C. (2014): "Why uncertainty matters - discounting under intertemporal risk aversion and ambiguity," *Economic Theoryl*, 56, 627 – 664.

Tritch, R. (2015): "President Obama could unmask big political donors," The New York Times, March 24.

Tsur, Y., and A. Zemel (2003): "Optimal transition to backstop substitutes for nonrenewable resources," *Journal of Economic Dynamics and Control*, 27, 551–572.

United Nations (2005): *Millennium Ecosystem Assessment.* United Nations, http://www.maweb.org/en/index.aspx.

van der Ploeg, R., and C. Withagen (2012): "Is there really a green paradox?," *Journal of Environmental Economics and Management*, 64, 342–363.

van der Werf, E., and C. Di Maria (2012): "Imperfect environmental policy and polluting emissions: the green paradox and beyond," *International Review of Environmental Resource Economicst*, 6, 153–194.

Vennemo, H., K. Aunan, F. Jingua, P. Hotedahl, H. Tao, and H. Seip (2006): "DOMESTIC ENVIRONMENTAL BENEFITS OF CHINAâĂŹS ENERGY-RELATED CDM POTENTIAL," *Climactic Change*, 75, 215–239.

Vissing-Jorgensen, A., and O. Attanasio (2003): "Stock market participation, intertemporal substitution and risk-aversion," *American Economic Review*, 93, 383 – 391.

V.Mcguire (2007): "Changes in water level and storage in the High Plains aquifer, predevelopment to 2005," U.S. Geological Survey Fact Sheet 2007-3029.

Weitzman, M. (1976a): "On the welfare significance of national product in a dynamic economy," *Quarterly Journal of Economics*, 90, 156–162.

——— (1976b): "The optimal development of resource pools," *Journal of Economic Theory*, 12, 351–364.

——— (1999): "Pricing the limits to growth from mineral depletion," *Quarterly Journal of Economics*, 114, 691– 706.

Weninger, Q. (1999): "Assessing efficiency gains from individual transferable quota: an application to the Mid-Atlantic surf clam and ocean quahog fishery," *American Journal of Agricultural Economics*, 80, 750–764.

Weninger, Q., and J. R. Waters (2003): "Economic benefits of management reform in the northern Gulf of Mexico reef fish fishery," *Journal of Environmental Economcis and Managment*, 46, 207–230.

Wilen, J., J. Cancion, and H. Uchida (2012): "The economics of Territorial Use Rights Fisheries," *Review of Environmental Economics and Policy*, 10.1093/reep/res011.

Winter, R. (2014): "Innovation and the Dynamic of Global Warming," *Journal of Environmental Economic and Management*, 68, 124–140.

World Bank (2014a): "The Little Green Data book 2014," Washington D.C. World Bank.

——— (2014b): "State and Trends of Carbon Pricing 2014," World Bank Group Climate Change.

Zhang, J., and M. D. Smith (2011): "Estmation of a generalized fishery model," *Review of Economics and Statistics*, 93, 690–99.