JEREMY GREENWOOD UNIVERSITY OF PENNSYLVANIA

RICARDO MARTO FEDERAL RESERVE BANK OF ST. LOUIS

NUMERICAL METHODS FOR MACROECONOMISTS

WITH JULIA AND MATLAB CODES

Contents

This primer will cover some of the numerical methods that are used in modern macroeconomics. You will learn how to:

1. Solve nonlinear equations via bisection and Newton's method;
2. Compute maximization problems by golden section search, discretization, and the particle swarm algorithm;
3. Simulate difference equations using the extended path and multiple shooting algorithms;
4. Differentiate and integrate functions numerically;
5. Conduct Monte Carlo simulations by drawing random variables;
6. Construct Markov chains;
7. Interpolate functions and smooth data;
8. Compute dynamic programming problems;
9. Solve for policy functions using the Coleman, endogenous grid, and parameterized expectation algorithms;
10. Solve the Aiyagari heterogeneous agent model with and without aggregate uncertainty.

This will be done while studying economic problems, such as the determination of labor supply, economic growth, and business cycle analysis. Calculus is an *integral* part of the primer and some elementary probability theory will be drawn upon. The MATLAB programming language will be used. It is time to move into the modern age and learn these techniques. Besides, using computers to solve economic models is fun. The primer is self contained so little prior knowledge is required.

If you like this primer, please cite it.

Comments and suggestions are welcome! This is a work in progress. Reports of errors, no matter how small, are greatly appreciated.

Contact Information: Jeremy Greenwood:
Telephone, (215) 898-1505; Fax, (215) 746-2947;
Email, NumericalMethods@jeremygreenwood.net—this email address is reserved *exclusively* for matters concerning the book.

1 Introduction

> There cannot be a language more universal and more simple, more free from errors and obscurities,....more worthy to express the invariable relations of natural things than mathematics. It interprets all phenomena by the same language, as if to attest the unity and simplicity of the plan of the universe, and to make still more evident that unchangeable order which presides over all natural causes.
> Joseph Fourier, *Analytical Theory of Heat*, 1822

> Many people have a passionate hatred of abstraction, chiefly, I think because of its intellectual difficulty; but as they do not wish to give this reason they invent all sorts of others that sound grand. They say that all reality is concrete, and that in making abstractions we are leaving out the essential. They say that all abstraction is falsification, and that as soon as you have left out any aspect of something actual you have exposed yourself to the risk of fallacy in arguing from its remaining aspects alone. Those who argue in this way are in fact concerned with matters quite other than those that concern science.
> Bertrand Russell, *The Scientific Outlook*, 1931.

Modern macroeconomics usually proceeds along the following path:

1. *Specifying people's tastes for goods and leisure*. This involves defining a utility function.

2. *Spelling out the technologies that individuals, firms, and governments employ to produce goods*. This could be a production function for firms, say using capital and labor. Sometimes a household production function is specified for a family that relates how much home goods are produced for a given amount of household labor and capital. Once in a while governments are assumed to produce goods as well, which also require capital and labor.

3. *Stipulating the structure of institutions and markets in the economy*. For example, do firms produce competitively or are they monopolistic in nature; what type of financial markets do households and firms have access to (say bonds, equities, and insurance); and what is the set of spending and tax instruments available to the government (for example, consumption, capital, and labor income taxes)?

4. *Solving the maximization problems for households and firms*. Usually households are assumed to maximize their utility and firms are taken to maximize their profits. This should be done before proceeding done the path to the next step.

5. *Imposing any equilibrium conditions*. For example, this might involve spelling out the capital, labor, and goods markets clearing conditions.

6. *Solving the government's maximization problem, if there is one*. Sometimes the government is taken as picking spending and taxes to maximize a social welfare function of some sort. Here the government is taken as a dominant player, whereby it knowingly influences the equilibrium in study. This explains its position in the steps. Other times the government's actions--spending and taxes--are just taken as given or are exogenous, as is assumed here.

7. *Studying the resulting economy*.

Of course, not all models have all of these ingredients. Some models don't have consumers, others don't have firms, and yet others exclude governments; there can be various combinations of these factors. Economies can be studied using old-fashioned pencil-and-paper techniques and/or modern numerical methods. Pencil-and-paper techniques are useful for developing propositions and theorems about economies. As model economies become more sophisticated it becomes increasingly difficult to develop propositions and theorems. Computers can be used to develop properties about economies, just as they are used in aerospace engineering to develop properties about air and spacecraft. Additionally, they allow for concrete quantitative predictions that are useful for policymakers, which are devoid in general mathematical analyzes.

Static economies can often be characterized as either solutions to nonlinear equation systems or as solutions to maximization problems in conjunction with an algorithm mimicking a Walrasian auctioneer. Chapter 2 studies labor supply in a static setting. This is done with and without government spending and taxation. It shows how this problem can be setup as the solution to a nonlinear equation. Various properties of the labor supply problem are established using pencil-and-paper techniques. The chapter then turns to discussing how this problem can be solved numerically using a nonlinear equation solver employing either the bisection algorithm or Newton's method. At the end of the chapter, a MATLAB program is presented that solves a monopolist's pricing problem in a static setting. Later on in Chapters 6 and 9, this example is made both dynamic and dynamic, stochastic.

Chapter 10 goes even one step further introducing heterogeneity across people into dynamic, stochastic settings.

In Chapter 3 the problem is recast as the solution to a maximization problem in conjunction with a Walrasian auctioneer. Three techniques are presented for maximizing a function; viz, golden-section search, discrete maximization, and particle swarm optimization. The chapter also discusses the concept of calibration. This involves choosing the parameter values for a model to maximize its fit with respect to a set of data targets. Two examples of calibrating models are presented. This first example focuses on the decrease in hours worked by males over the course of the 20th century. The second example discusses the rise in the premarital sexual activity over the last century.

As an example of studying taxation in a static economy, the Prescott (2004) study on "Why Do Americans Work So Much More Than Europeans?" is discussed in Chapter 4. The chapter also illustrates how a parameter value can be chosen to maximize the fit of a model; this is an exercise in calibration. Since macro model's are often calibrated to the national income and product accounts, as Prescott does, a brief discussion of these are presented. The national income and product accounts are must knowledge for macroeconomists.

The discussion then moves on in Chapter 6 to the solution of deterministic dynamic models. This is cast within the context of the neoclassical growth model. To begin with, the deterministic dynamics of the model are completely characterized using pencil-and-paper techniques. It is shown how the solution to this model can be formulated as a nonlinear difference equation. While doing this, the notions of dynamic programming and the value function are introduced. Properties of the value function for the neoclassical growth model are derived. All of this is done in a heuristic (non-rigorous) way so as not to obscure the beauty of dynamic programming by intimidating readers.

Next, three techniques for numerically solving this nonlinear difference equation are presented; viz, the extended path method, linearization, and multiple shooting. Multiple shooting picks one of the initial conditions for the nonlinear difference equation so that the economy ends up in its steady state after some extended period of time. The extended path method conjectures a path for expectations about how the economy will evolve. It then solves the economy given this conjectured path and computes a revised path for the expectations. The procedure is then repeated. When the path for expectations and the path for the actual economy converge a perfect foresight path has been found. Linearization refers to a method where the nonlinear difference equation describing an economy's dynamics is approximated by a linear difference equation. The chapter illustrates how this is done. Finally, the monopolist's pricing problem, introduced in Chapter 2, is returned to. The problem is now made dynamic. A MATLAB program is provided that solves this problem using the three solution techniques discussed. As an example of deterministic dynamics, "Malthus to Solow" by Hansen and Prescott (2002) is presented in Chapter 7.

Chapter 9 deals with stochastic dynamics. The discussion is centered around the stochastic growth model, which is widely used in business cycle analysis. Three numerical techniques are presented for solving dynamic stochastic economies; namely, dynamic programming, linearization, and policy-function iteration. Three methods for policy-function iteration are presented: the Coleman (1991) algorithm, the Carroll (2006) endogenous grid method, and den Haan and Marcet (1990) parameterized expectations. As an illustration of these techniques, the monopolist's dynamic pricing problem presented in Chapter 6 is now made stochastic. A MATLAB program illustrating how to solve this model is presented.

The famous Aiyagari (1994) model is presented in Chapter 10. This was one of the first papers to extend the standard representative agent model to a world with heterogeneous agents. In the original Aiyagari model there was no aggregate uncertainty. Individuals face idiosyncratic randomness in their labor income. They also faced a borrowing constraint that limited their ability to insure against the risk in their labor income. Aiyagari showed how a distribution of wealth emerges across people. Boppart et al. (2018) illustrate how the Aiyagari model can be extended to incorporate aggregate uncertainty. Their methodology for doing this is discussed. To illuminate the workings of Aiyagari-type models the stochastic, dynamic monopolist's problem presented in Chapter 9 is extended to cover the case of monopolistic competition. Here a firm's demand for its product depends upon the distribution of output across other firms in the economy. The greater the sales by other firms, ceteris paribus, the lower is the demand for the firm's own product at any given price.

Some numerical approximations that are useful for solving macroeconomic models numerically, especially stochastic ones are covered in Chapter 8. The chapter starts off discussing numerical derivatives. Two methods are covered here: the standard method and complex step differentiation. These days computers can also differentiate analytical expressions symbolically, so this is touched upon. Next, the chapter turns to the classical method for numerical integration. As an example of this technique, the consumer surplus for computers is calculated. The chapter then moves on to random number generation. This topic is illustrated using an early business cycle model à la Slutsky (1937). Random number generation leads naturally to the subject of Monte Carlo integration. As an example of this, the chapter visits the welfare cost of business cycles as advanced by Lucas (1987). The concept of a Markov chain is also presented. The usefulness of Markov chains is illustrated using two examples. The first illustration constructs a Markov chain for unemployment and uses this to estimate job finding and separation rates. The second illustration is the Mehra and Prescott (1985) study on equity premium. A method for approximating an AR1 process by a Markov chain is discussed. The last topic in the chapter is the approximation of functions. Often in macroeconomics one wants to compute some functions for which there are no known analytical solutions, such as policy functions or value functions. Three methods are discussed for approximating functions: piecewise linear interpolation, cubic spline interpolation, and radial basis function interpolation. Cubic spline interpolation is a very flexible technique. This is shown by mimicking an artist's sketch of a face using cubic splines. This is also a natural point to introduce the Hodrick-Prescott filter.

Graphing data and the results from models is an important part of macroeconomics. Statistical graphing was introduced in the 18th century by an economist, William Playfair. Chapter 5 discusses some basic principals for graphing. Some of Playfair's beautiful graphs are reproduced. MATLAB programs for three Playfair-style graphs are provided.

The book is self contained. Chapter B provides an introduction to MATLAB. The elementary mathematics used in the book are reviewed in Chapter A. A legend for some of the notation used in the book is also presented here. It is important for economists in the modern era to be able to move fluidly between economics, computing, and mathematics. There is synergy between these three branches of knowledge. Mathematics forces clarity of thought and is necessary for setting up economic models on computers. Computers are needed for solving complicated economic models and for providing concrete predictions. Writing computer code also fosters a better understanding of mathematical concepts since the math has to be operationalized in a practical sense. Most important of all is a firm understanding of economics. First, one needs to know what an interesting economic problem is. Second, it's crucial for understanding the economic intuition that is embedded in mathematical formulations--the math should speak to you. Third, an adage in computer science is "garbage in, garbage out." Having a sense of what to input into an economic model and what to expect out is very important.

Last, this genre of economics is often called *Quantitative Theory*. It both complements and overlaps with conventional econometrics. Nonstructural econometrics formulates *statistical* models. There may or may not be an *economic* model that gives rise to the functional forms estimated in nonstructural econometrics. Statistical models are very useful for characterizing facts in the data. The interaction between measurement and theory is bidirectional. One should remember Tjalling C. Koopmans's (1910-1985) warning about measurement without theory.[1] Empirical findings motivate theory and theory sheds light on what ideas to test and guides empirical formulations. Results from nonstructural econometric models (e.g., regression coefficients) can be used to calibrate simulated models. This is called indirect inference. Structural econometric models are a close cousin to calibrated ones. On the one hand, they use formal statistical analysis to evaluate the model. On the other hand, due to the added difficulties of estimation, they are often partial equilibrium in nature. Also, sometimes minimum distance estimation procedures are used to calibrate simulated models so that quantitative theory and structural econometrics overlap.

[1] He critiqued empirical work in *Koopmans (1947)* saying: 'The various choices as to what to "look for," what economic phenomena to observe, and what measure to define and compute, are made with a minimum of assistance from theoretical conceptions or hypothesis regarding the nature of the economic processes by which the variables studied are generated (p. 161).' This view is reverberated by Nobel prize winner James J. Heckman who in a 2022 interview stated "Blind empiricism unguided by theoretical frameworks for interpreting facts leads to nowhere."

2 Nonlinear Equations

2.1 Introduction

In modern macroeconomics an economy can often be described as the solution for h to a nonlinear equation of the following form:

$$Z(h) = 0. \tag{2.1.1}$$

Here h could be a single variable or a vector of variables and likewise Z could be a single or vector valued function. The function Z should have the same dimension as the variable h; i.e., you need the same number of equations as unknowns. A solution to this equation is called a zero of Z or the root of $Z(h) = 0$. Solving such equations is the subject of this chapter. Embodied in Z may be the tastes and technology of the economy, the tax and spending policies of the government, the upshot of individuals' and firms' choice problems, and market-clearing conditions. In fact, modern macroeconomics tries to specify the economy at the granular level needed to address the question of interest. As government policy or technology change, so will the form of the function $Z(h)$. By modeling things at the granular level such shifts in the function will be captured.

Two methods are presented in Section 2.6 for solving the above equation, namely the bisection method and Newton's method. The chapter starts out in Section 2.2 with a consumption-leisure choice problem. The discussion follows the path outlined in Chapter 1 for modern macroeconomics. The impact of shifts in government spending, taxes, and wages are broken down into income and substitution effects along the lines of Sir John R. Hicks (1904-1989) and Eugen Slutsky (1880-1948). The income effect associated with a government fiscal policy depends crucially on how the tax revenue raised is used. In particular it depends on whether the revenue is used for either transfer payments or government spending on goods and services. Furthermore, if the revenue is used to provide goods and services, does this government spending substitute for private expenditure or not? The analysis is then cast into a general equilibrium setting where the wage rate and rental rate on capital are determined endogenously. To measure the change in economic welfare associated with a shift in government policy, the Hicks (1941) notions of compensating and equivalent variations are introduced. These notions are depicted using the Lucas (1987) calculation of the welfare benefit/costs from changing an economy's growth rate. It is shown how the general equilibrium solution to the consumption-leisure choice problem in economy with taxes and government spending can be setup as a nonlinear equation that has the form of equation (2.1.1). Pseudo code illustrating the two techniques for solving nonlinear equations is presented in Section 2.7 for a monopolist's pricing problem. Last, the chapter ends with a discussion of Leonfief's input-output framework, which represents a simple general equilibrium model composed of a system of linear equations.

2.2 Consumption-leisure Choice

2.2.1 Utility Functions

Tastes are specified by a utility function. Let

$$u = U(c)$$

represent the utility function for consumption. It gives the level of happiness, u, that person realizes if they consume the amount, c. Utility is an *ordinal* concept, not an cardinal one. It specifies how different consumption levels are ranked. The precise numbers assigned to particular consumption levels are meaningless. It's important to remember this when conducting welfare experiments, as will be discussed in Section 2.5. Some typical properties imposed jointly on a utility function, or on the ordinal ranking, are:

1. $U : \mathcal{R}_+ \to \mathcal{R}$ (so that a utility function maps the positive reals into the reals). Consumption must always be nonnegative, but utility can be negative.

2. U is strictly increasing so that $U_1 \equiv dU/dc > 0$. More of a good is better than less of it. Marginal utility is positive. Even if utility is negative it will be increasing in consumption, because marginal utility is positive. Now suppose one added a constant to the utility function. The utility value connected with different levels of consumption would change by the added constant. In fact, utility could be made arbitrarily large or small by maniputing this constant rendering the exact value meaningless. But, utility would still be strictly increasing with exactly the same first derivative and the original ranking across different levels of consumption is preserved. This illustrates the ordinal nature of utility.

3. U is strictly concave so that $U_{11} \equiv d^2U/dc^2 < 0$. Marginal utility decreases as consumption increases. Each extra increment of consumption is worth less to the consumer. This is called diminishing marginal utility. Again, adding a constant to the utility function does not change the second derivative. So the exact values assigned to utility arising from a particular level of consumption are not meaningful. As will be seen, the assumption that utility functions are strictly concave plays an important role in economics.

Example 2.1. (Common utility functions) Here are some utility functions that are commonly used in macroeconomics. They satisfy the above properties.

$U(c) = \ln c$ (logarithmic);

$U(c) = c^{1-\rho}/(1-\rho) - 1/(1-\rho)$, with $\rho \geq 0$ (isoelastic);

$U(c) = -e^{-\gamma c}$, with $\gamma > 0$ (exponential);

$U(c) = \alpha c - \beta c^2/2$, with $\alpha, \beta > 0$ and for $c < \alpha/\beta$ (quadratic).

These utility functions are illustrated in Figure 2.2.1. The logarithmic utility returns a negative value for $c < 1$ and positive one for $c > 1$. Observe that $U_1 = d\ln c/dc = 1/c > 0$ and $U_{11} \equiv d^2 \ln c/dc^2 = -1/c^2 < 0$. Utility can also be positive or negative with an isoelastic utility. Here $U_1 = c^{-\rho} > 0$ and $U_{11} = -\rho c^{-\rho-1} < 0$. The isoelastic utility function is also called a constant-relative-risk-aversion (CRRA) utility function. Here $1/\rho$ represents the elasticity of intertemporal substitution, which is defined in Chapter 6. The elasticity of intertemporal substitution controls the responsiveness of consumption to interest rate changes. With this utility function ρ also represents the coefficient of relative risk aversion, as will be discussed in Chapter 8. This governs an individual's willingness to invest in risky assets. So, ρ plays two roles, which may cause problems in some applications. Often the constant term, $-1/(1-\rho)$, is dropped. When this included, the isoelastic utility function converges to the logarithmic one as ρ approaches 1.

Utility is always negative with the exponential utility function. So, the sign of utility is not important. The important thing is that as c rises so does utility and this increase in utility decreases with the level of c. The quadratic utility function is $\cap$ shaped. As can be seen, $U_1 = \alpha - \beta c \gtreqless 0$ depending on whether $c \lesseqgtr \alpha/\beta$, Therefore, this function rises or falls depending on the value of c. Only the upward portion of the $\cap$ is valid. The peak of the $\cap$ occurs at $c = \alpha/\beta$. This explains the restriction imposed on c. The quadratic utility function is still strictly concave because $U_{11} = -\beta < 0$. Sometimes quadratic utility functions are used in numerical work to approximate nonquadratic ones, as is done in Chapter 6.

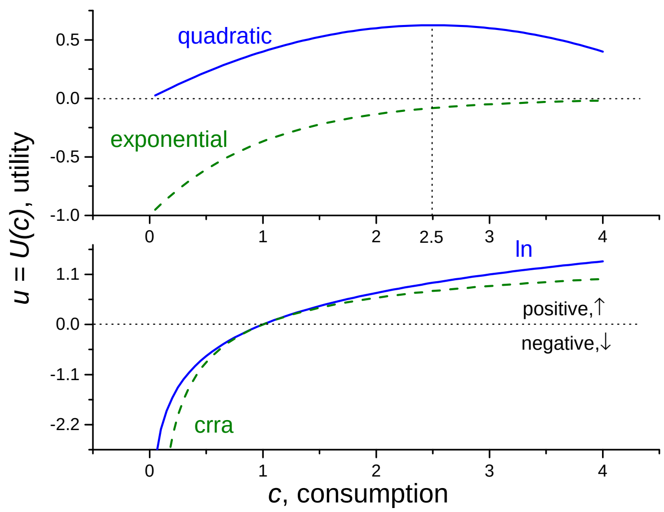


Figure 2.2.1: Utility functions: logarithmic (ln); crra ($\rho = 1.5$); exponential ($\gamma = 1$); quadratic ($\alpha = 0.5, \beta = 0.2$). The exponential utility function always returns a negative value for utility, u. The crra and ln utility functions can yield both negative and positive values for utility. When $\rho > 1$ the crra utility function is more concave than the ln one and less so when $\rho < 1$. The quadratic utility function declines when $c > \alpha/\beta$. Hence, it is only good for $c < \alpha/\beta$. The peak of the quadratic utility function occurs at $c = \alpha/\beta = 2.5$.

Likewise, let

$$v = V(1-h)$$

represent the utility function for leisure. Here it is presumed that an individual has one unit of time that can be split between working and leisure. This utility function returns the level of happiness, v, that a worker realizes if he spends the proportion of his time h working so that he enjoys the fraction $1-h$ in leisure. Leisure can be thought of as a good, just as consumption is, so the properties for V are the same as those imposed on U.

2.2.2 *A Static Consumer/Worker's Decision Problem*

Suppose that a person works on a spot market for labor. They have one unit of time that they can split between working, h, and leisure, $1-h$. The wage rate for a unit of labor is w. Let a denote the worker's level of assets, which is exogenously specified for the moment. This will be used to calculate the effect of wealth, a, on labor supply, h. Later in this chapter a will be connected with the rental income that accrues from a fixed amount of capital or land. Additionally, in Chapter 6 the rental income that a person will earn from their ownership of capital will be the outcome of a consumer/worker's consumption/savings problem. The worker's static maximization problem is

$$\max_{c,h}\{U(c) + V(1-h)\},$$

subject to the budget constraint

$$c = wh + a.$$

William Stanley Jevons (1835-1882) was the first to introduce the notion of utility maximization in his 1871 book *The Theory of Political Economy*, where he represented utility as a mathematical function. He held a chair at University College, London. While on a vacation, he drowned while swimming, most probably from a heart attack or stroke.

The lefthand is the person's expenditure on consumption, c. The righthand represents the resources at their disposal, which derive from their labor income, wh, and their wealth, a. By substituting the budget constraint into the objective function to solve out for consumption, c, a maximization problem in just hours worked, h, can be obtained. The revised maximization problem is

$$\max_{h}\{U(wh+a)+V(1-h)\},$$

The objective function for this problem is shown in Figure 2.2.2.

The slope of the objective function is zero at a maximum. (See the Mathematical Appendix in Chapter A for an elementary exposition of maximization problems.) This corresponds to setting

$$U_1(wh+a)w - V_1(1-h) = 0,$$

which is the first-order condition to the single-variable maximization problem. When the objective function is strictly concave in h, this first-order condition is both a necessary and sufficient condition for a maximum in h to attain. This equation has the form of (2.1.1), which can be seen by making the following definition for $Z(h)$:

$$Z(h) \equiv U_1(wh+a)w - V_1(1-h) = 0.$$

The above describes one equation in the one unknown endogenous variable, h, where a and w are exogenous variables. By the implicit function theorem, the solution for h to this equation can be written as $h = H(a,w)$. The function H is the person's *decision rule*. It gives their optimal choice for labor effort, h, as a function of the exogenous variables a and w. (The implicit function theorem is presented in Chapter A.)

The above first-order condition can be rewritten as

$$\underbrace{U_1(wh+a)w}_{\text{Marginal benefit from working}} = \underbrace{V_1(1-h)}_{\text{Marginal cost of working}}. \tag{2.2.1}$$

The solution is portrayed in Figure 2.2.3. The righthand side is increasing in h, because V_1 is decreasing in $1-h$. This results from the assumption that the utility function for leisure is strictly concave or that leisure exhibits diminishing marginal utility; i.e., $V_{11} = dV_1(1-h)/d(1-h) < 0$. The righthand side represents the marginal cost of working; therefore, the marginal cost of working is increasing in hours worked, h. The lefthand side is decreasing in h, because U_1 is decreasing in c, and hence $wh+a$, due to the fact that utility is strictly concave in consumption ($U_{11} = dU_1(c)/dc < 0$). The lefthand side portrays the marginal benefit of working; hence, the marginal benefit of working decreases in labor effort, h. This arises because the utility function for

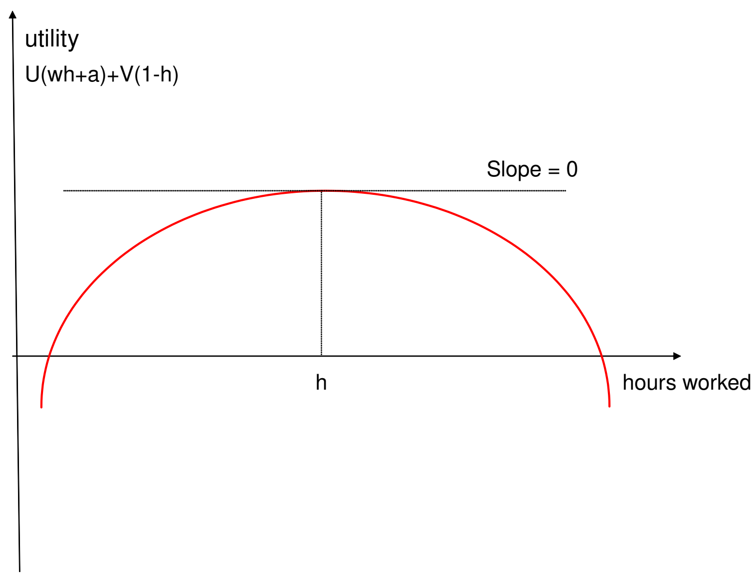


Figure 2.2.2: The representative consumer/worker's objective function. The level of hours worked, h, that maximizes utility occurs where the slope of utility function is zero, or where $Z(h) = U_1(wh+a)w - V_1(1-h) = 0$.

consumption is strictly concave in h or because consumption displays diminishing marginal utility. Your consumption increases as you work more, but the value of the extra consumption is subject to diminishing marginal utility.

In principal a corner solution can occur–the mathematics of corner solutions is discussed in Chapter A. For example, the person would not want to work ($h = 0$) when

$$\underbrace{U_1(a)w}_{\text{Marginal benefit from working}} < \underbrace{V_1(1)}_{\text{Marginal cost of working}}.$$

Here the marginal benefit of working (at $h = 0$) is less than its marginal cost. It's likely that there exists a large enough value for a such that the corner solution holds for sure. For example, think about the situation where $\lim_{a\to\infty} U_1(a) = 0$. The person is so wealthy that an extra unit of consumption is worthless to them. In this case (not shown) the marginal cost curve in Figure 2.2.3 lies above the marginal benefit curve at $h = 0$, which is flat at zero for all values of h.

The impact that shifts in wealth, a, and wages, w, have on labor supply, h, are now analyzed. The notions of income and substitution effects, as advanced by Hicks (1939) and Slutsky (1915), come into play here.

Sir John Hicks (1904-1989) was a British economist. Hicks won the Nobel Prize in Economics in 1972. He brought many important ideas into economics: income and substitution effects, compensating and equivalent variations, and the IS-LM model.

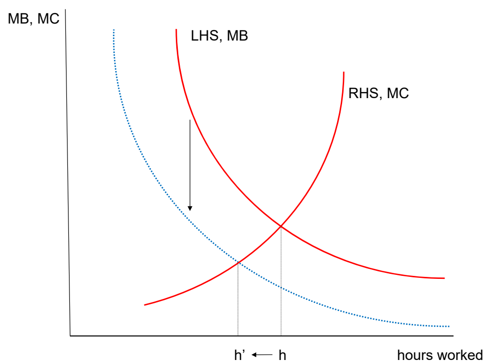


Figure 2.2.3: Consumption-leisure choice. The optimal level of hours worked, h, occurs where the marginal benefit from working, MB, equals its marginal cost, MC–compare with equation (2.2.1). An increase in wealth, a, causes the MB curve to shift down, which results in a decline in hours worked from h to h'. This illustrates the wealth effect on labor supply.

The impact of an increase in wealth, a

Suppose the worker is wealthier; i.e., increase a. By examining the righthand side of equation (2.2.1), it is clear that the marginal cost curve in Figure 2.2.3 is not a function of a. From the lefthand side of (2.2.1) it can be seen that the marginal benefit curve is. In particular, for any given level of h, an increase in a will cause $c = wh + a$ to rise. Hence $U_1(wh + a)$ falls, due to diminishing marginal utility or the fact that the utility function is strictly concave. Thus, the marginal benefit curve shifts down. This results in a fall in labor supply, h–see Figure 2.2.3. When the person gets wealthier, they would like to spread the windfall across both consumption and leisure. They won't use all of the windfall for consumption because the marginal utility of consumption is declining so that each extra unit of consumption is worth less and less. Hence some of the gain in wealth should be directed toward leisure.

To obtain the mathematical transliteration of the graphical analysis take the total differential of (2.2.1) with respect to a and h. This gives

$$U_{11}(wh + a)w^2 dh + U_{11}(wh + a)w da = -V_{11}(1 - h)dh,$$

which yields

$$\frac{dh}{da} = \frac{-U_{11}(wh + a)w}{U_{11}(wh + a)w^2 + V_{11}(1 - h)} < 0. \tag{2.2.2}$$

The concepts of total differentials and total derivatives are reviewed in Chapter A. The sign of the above expression results from the fact that $U_{11}(wh + a)$ and $V_{11}(1 - h)$ are both negative because the utility functions for goods and leisure are assumed to be strictly concave. So, the result that hours worked declines with wealth hinges on the assumption that the utility functions for consumption and leisure are strictly concave. By multiplying both sides of the above expression by w it is easy to deduce that $0 < d[w(1-h)]/da = -wdh/da < 1$. Therefore, the consumer/worker is using part of his increased wealth to increase spending on both consumption and leisure. As will be seen, the term on the righthand side of (2.2.2) is closely connected with the income effect from a rise in wages.

The impact of a rise in wages, w

Suppose that wages rise. It's hard to tell if the marginal benefit curve will shift to the right or left. There are two opposing forces, as can be seen by examining the lefthand side of (2.2.1). First, holding marginal utility constant, or $U_1(wh+a)$, an increase in w raises the marginal benefit from working. In response, hours worked will rise. This is the substitution effect from the rise in wages. Second, holding h fixed, an increase in w operates to reduce the marginal utility of consumption. Hence, on this account, the marginal benefit from working will drop. Hours worked will fall and hence leisure rise. This represents the income effect from an increase in wages. Mathematically one finds, by taking the total differential of (2.2.1) with respect to h and w, that

$$U_{11}(wh+a)w^2dh + U_{11}(wh+a)hwdw + U_1(wh+a)dw = -V_{11}(1-h)dh, \tag{2.2.3}$$

$$\begin{aligned}\frac{dh}{dw} &= \underbrace{\frac{-U_{11}(wh+a)wh}{U_{11}(wh+a)w^2 + V_{11}(1-h)}}_{\text{Income Effect, }<0} + \underbrace{\frac{-U_1(wh+a)}{U_{11}(wh+a)w^2 + V_{11}(1-h)}}_{\text{Substitution Effect, }>0} \gtreqless 0 \\ &= h\frac{dh}{da} + \frac{-U_1(wh+a)}{U_{11}(wh+a)w^2 + V_{11}(1-h)} \gtreqless 0, \text{ using (2.2.2).}\end{aligned}$$

Thus, the effect is ambiguous depending on the relative size of the income and substitution effects. The income effect is given by the first term. The size of the income effect is proportional to amount of work that the person does, h. If he did little work ($h \simeq 0$), then he would not gain much in income from an increase in the wage rate.

Example 2.2. (Logarithmic utility) Let $U(c) = \theta \ln c$ and $V(1-h) = (1-\theta)\ln(1-h)$ where $0 < \theta < 1$. This is the most commonly used functional form for utility in macroeconomics. Then, the above first-order condition appears as

$$\underbrace{\theta\frac{1}{wh+a} \times w}_{\text{MB}} = \underbrace{(1-\theta)\frac{1}{1-h}}_{\text{MC}}.$$

Cross multiplying and solving for h yields:

$$h = \frac{\theta w - (1-\theta)a}{w} = \theta - (1-\theta)\frac{a}{w},$$

at least when there is an interior solution. Note that it is possible for $h = 0$, which occurs when the above equation returns a negative solution for h. A negative value for hours worked, h, is invalid; the lowest it can be is zero. It follows that

$$\frac{dh}{dw} = (1-\theta)\frac{a}{w^2} \gtreqless 0.$$

There are three cases to consider.
(i) If $a > 0$, then labor supply, h, is increasing in wages, w. Hence, the substitution effect is larger than the income effect. As an illustration of this case, think of a married household where the husband works a fixed work week and a is his income. Here, w represents the wife's wage. Hence, in a married household a rise in the wife's wage, w, would cause her to work more. A married woman may not work ($h = 0$), if her husband earns enough. This occurs when the marginal benefit from working, MB, is less than the marginal cost, MC, when evaluating (2.2.1) at $h = 0$.
(ii) If $a = 0$, then a change in wages will have no impact on labor supply, because the income and substitution effects from a rise in wages would exactly cancel out. Also, observe that if assets were proportional to wages, so that $a = \psi w$ for any $\psi \geq 0$, then a rise in wages, w, would have no impact on hours worked, h; again, the income and substitution effects exactly cancel out. This property is used in balanced growth models, where wealth rises in tandem with wages.
(iii) If $a < 0$, then an increase in wages will reduce hours worked. Here the income effect dominates the substitution effect.
Taking stock of things, the strength of the income effect is decreasing in assets, a. This makes sense. An extra dollar must be worth more to a poor person vis à vis a rich one.

This example is solved *analytically* using the computer in Section 8.3 of Chapter 8.

Example 2.3. (Subsistence level of consumption) Rewrite the utility function as $U(c) = \theta \ln(c - \mathfrak{c})$ and give the person the budget constraint $c = wh$; i.e., assume that the person has no assets, a. Here $\mathfrak{c}$ can be interpreted as a subsistence level of consumption. The solution to this case can be obtained from the previous example by letting $\mathfrak{c} = -a > 0$ (so that the solution will operate in the same manner as assuming that $a < 0$). Therefore, in this illustration the solution for h, associated with the consumer/worker's maximization problem, is the same as before: $h = \theta + (1-\theta)\mathfrak{c}/w$. The higher the subsistence level of consumption, or $\mathfrak{c}$, is, the harder the person will work. As wages rise the person will work less.

Example 2.4. (Zero-income effect utility function) Let utility be given by $W(c,h) = \ln(c - h^{1+\eta}/(1+\eta))$. The worker's problem is

$$\max_h W(\underbrace{wh + a}_{c}, h).$$

The first-order condition for labor reads

$$\underbrace{W_1(c,h) \times w}_{\text{MB}} = \underbrace{-W_2(c,h)}_{\text{MC}}.$$

Plugging in the above functional form gives

$$\frac{1}{c - h^{1+\eta}/(1+\eta)} \times w = \frac{h^\eta}{c - h^{1+\eta}/(1+\eta)},$$

so that

$$h = w^{1/\eta}.$$

Labor supply, h, is increasing in wages, w. Here $1/\eta$ gives the elasticity of labor supply with respect to wages. Using the above formula it is easy to calculate that

$$\frac{w}{h}\frac{dh}{dw} = \frac{1}{\eta},$$

where the lefthand side gives the percentage change in hours worked in response to a percentage change in wages. This utility function has no income effect. Note the absence of a in the solution for h. This utility function is often used in business cycle modeling because of its simple solution for h. It will be returned to in Chapter 6.

2.3 *Government Spending and Taxation*

How do government spending and labor income taxation affect hours worked? This will be explored now. It will be discovered that the effect depends upon how the revenue from the taxation is used. This governs the income effect associated with a government spending-cum-tax plan. Specifically, the impact of taxation will differ according to whether the goverment uses the revenue for:

1. Lump-sum transfer payments. Lump-sum transfer payments imply that taxation has a zero income effect because all revenue is rebated back to the consumer/worker.

2. Government spending. There are two cases to consider here:

   (a) Government spending does not substitute for private consumption spending. Here there will be a negative income effect associated with the government spending because the government is drawing resources out of the economy that will reduce private consumption spending. This is true even if the government spending is valued by consumers.

(b) Government spending substitutes for private consumption spending. The negative income effect will be mitigated to the extent that the spending substitutes for reduced private consumption.

Suppose that there is a government in the economy that taxes labor income at the rate τ. It uses the revenue raised from labor income taxes to finance government spending on goods and services, g, and to provide lump-sum transfer payments, λ. The strength of the impact of taxation on labor supply depends crucially on how the revenue raised is used.

The government's budget constraint appears as

$$\underbrace{g+\lambda}_{\text{disbursements}} = \underbrace{\tau wh}_{\text{receipts}}. \tag{2.3.1}$$

To calculate the effect of taxation on labor supply follow the path outlined in Chapter 1:

1. Solve the individual's consumption-leisure choice with labor taxation and transfer payments to obtain the individual's first-order condition.

2. Impose any equilibrium conditions and the government's budget constraint on the first-order condition.

Incorporating any equilibrium conditions and/or government's budget constraint into the individual's consumption-leisure choice before solving the person's problem leads to a *fundamental error* in economics, which is discussed later on.

2.3.1 *Step 1, The worker's problem with taxation*

Start with the case where the government taxes labor income at the rate τ, provides lump-sum transfers in the amount, λ, and spends g. The consumer/worker does not value government spending here. Valued government spending will be discussed later. The worker's problem is now

$$\max_{c,h}\{U(c)+V(1-h)\},$$

subject to his budget constraint

$$c=(1-\tau)wh+a+\lambda.$$

The worker's after-tax labor income is $(1-\tau)wh$ when he works the amount h. The level of lump-sum transfer payments, λ, is unrelated to the individual's work effort. This provides an additional source of funds for consumption spending, c. After using the budget constraint to solve out for c in the utility function, it is easy to deduce that the first-order condition for h is

$$Z(h) \equiv U_1\Big((1-\tau)wh + a + \lambda\Big) \times (1-\tau)w - V_1(1-h) = 0,$$

which can be rewritten as

$$U_1\Big(\underbrace{(1-\tau)wh + a + \lambda}_{\text{IE}}\Big) \times \underbrace{(1-\tau)w}_{\text{SE}} = V_1(1-h). \tag{2.3.2}$$

Again, observe that this is one equation in one unknown. A value of h is sought that sets $Z(h) = 0$. Represent the solution by $h = H(w, \tau, \lambda, a)$. Will income taxation raise or reduce work effort? On the one hand, a unit of labor now earns the after-tax wage rate, $(1-\tau)w$. This creates a disincentive to work, at least when consumption $c = (1-\tau)wh + a + \lambda$ is held fixed. This is the substitution effect, SE, from taxation. On the other hand, the person's income, $(1-\tau)wh + a + \lambda$, and hence consumption will be lower, for a given level of hours worked, h. This operates to make the person work harder, other things equal. This is the income effect, IE.

2.3.2 *Step 2, Impose the government's budget constraint*

Proceed now to Step 2. By using the government's budget constraint (2.3.1), to solve out for lump-sum transfers, λ, in the above first-order condition (2.3.2), one obtains

$$U_1(wh + a - g)(1-\tau)w = V_1(1-h). \tag{2.3.3}$$

It is easy to see intuitively how the government can affect the worker. First, the presence of taxation distorts the person's labor supply decision, as shown by the $(1-\tau)$ term. This works as a negative substitution effect. Second, the government takes away some of the economy's resources, as reflected by the g term. This operates as a negative income effect–the situation by portrayed in Figure 2.3.1. If there is no government spending, such as when all revenue is rebated back as transfer payments, there will be no income effect associated with the taxation. The case of transfers is discussed next.

2.3.3 *All taxes rebated back as lump-sum transfers*

Here,

$$\tau wh = \lambda \text{ and } g = 0.$$

The worker's consumption will be

$$c = (1-\tau)wh + a + \lambda = wh + a.$$

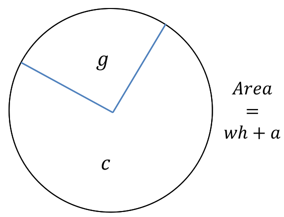


Figure 2.3.1: The size of the economy is $wh + a$, or the area of the circle. Out of this pie the government takes the slice g leaving c for private consumption. The size of the pie is endogenous, depending on labor supply, h. Labor supply is a function of the level of government spending, g, and the rate of labor income taxation, τ.

Plug this into the first-order condition (Step 2) to get

$$U_1(wh + a)\underbrace{(1 - \tau)w}_{\text{SE}} = V_1(1 - h), \tag{2.3.4}$$

or equivalently just set $g = 0$ in (2.3.3). The way taxes enter the above expression suggests that only the substitution effect will be operational.

Now an increase in taxes reduces the marginal benefit from working. Hence, labor supply falls. Parroting the above exercises one gets

$$\frac{dh}{d\tau} = \frac{U_1(wh + a)w}{U_{11}(wh + a)(1 - \tau)w^2 + V_{11}(1 - h)} < 0.$$

This is just the substitution effect–compare with the second term in (2.2.3). In terms of Figure 2.2.3 an increase in taxes will shift the MB curve down for any given value of h. This transpires because the after-tax wage rate, $(1 - \tau)w$, drops. The MC curve remains fixed.

Remark 2.1. (The meaning of representative agent models) The construct of a representative agent is a stand-in device for millions of identical individuals each maximizing their own welfare while taking the actions of other parties in the economy as given. A huge blunder for a macroeconomist to make is to substitute the government's budget constraint into the consumer/worker's one *before* the maximization is done. If this is done, then (2.2.1) will reappear instead of (2.3.2). Hence, there would be no apparent effect of taxes on labor supply. To understand the mistake suppose that there are n identical agents in the economy. Let the representative agent choose labor supply in the amount h. Suppose that the other $n - 1$ people pick $\mathbf{h}$. Of course in equilibrium $h = \mathbf{h}$, because they are all the same. The government's budget constraint can be written as

$$\tau wh + (n - 1)\tau w\mathbf{h} = n\lambda,$$

or

$$\lambda = \tau wh/n + (n - 1)\tau w\mathbf{h}/n.$$

The person knows if he works one unit of time more, and no one else does, his transfers will increase by $\tau w/n$. The government's revenue from this person's extra work effort is shared by all people equally. Thus, the other $n-1$ individuals in the economy will also each receive $\tau w/n$ in extra transfer payments, even though they haven't changed their work effort. The person's maximization problem is now

$$\max_h \Big\{ U\Big((1-\tau)wh + a + \tau wh/n + (n-1)\tau w\mathbf{h}/n\Big) + V(1-h)\Big\},$$

where λ has been solved out for in his budget constraint. The individual cannot tell his neighbor what to do, so he must take $\mathbf{h}$ as given in this maximization problem. The agent's first-order condition is

$$U_1\Big((1-\tau)wh + a + \tau wh/n + (n-1)\tau w\mathbf{h}/n\Big)(1-\tau+\tau/n)w = V_1(1-h).$$

The person realizes that if he works an extra hour, then his take will be $(1-\tau+\tau/n)w$, because now he will get $(\tau/n)w$ rebated back from the taxes he pays. Now set $\mathbf{h} = h$ because all individuals will be the same in competitive equilibrium. The above condition then appears as

$$U_1(wh+a)(1-\tau+\tau/n)w = V_1(1-h).$$

Clearly, as $n \to \infty$ this converges to (2.3.4). When $n = 1$ it looks like (2.2.1). Thus, the mistake is treating the representative agent as being the single person in the economy instead of as representing millions of identical individuals, each acting on their own, while taking as given other people's actions.

2.3.4 Non-Valued Government Spending

Consider the case where all tax revenue is used to finance government spending on goods and services, g. Since the consumer/worker does not value the government spending, g does not enter his utility function, in contrast to the case of valued government spending discussed below. Since there are no lump-sum transfers, $\lambda = 0$ so that

$$\tau wh = g.$$

Plugging this revised government budget constraint into the first-order condition (2.3.3) yields

$$U_1\Big(\underbrace{(1-\tau)wh + a}_{\text{IE}}\Big)\underbrace{(1-\tau)w}_{\text{SE}} = V_1(1-h). \tag{2.3.5}$$

Now an increase in taxes has an ambiguous impact on the marginal benefit from working, since a rise in taxation has both an income and substitution effect. Parroting the above exercises yields

$$\frac{dh}{d\tau} = \underbrace{\frac{U_{11}\Big((1-\tau)wh+a\Big)(1-\tau)w^2h}{U_{11}\Big((1-\tau)wh+a\Big)(1-\tau)^2w^2+V_{11}(1-h)}}_{\text{income effect, }>0} + \underbrace{\frac{U_1\Big((1-\tau)wh+a\Big)w}{U_{11}\Big((1-\tau)wh+a\Big)(1-\tau)^2w^2+V_{11}(1-h)}}_{\text{substitution effect, }<0} \lesseqgtr 0.$$

Whether hours worked will fall or rise depends on whether the substitution or income effect dominates.[1]

[1] It is easy to calculate that

$$\frac{dh}{da} = -\frac{U_{11}\big((1-\tau)wh+a\big)(1-\tau)w}{U_{11}\big((1-\tau)wh+a\big)(1-\tau)^2w^2+V_{11}(1-h)} < 0.$$

From this it is clear that the first term in the above expression is the income effect; i.e., first term can be rewritten as $-wh \times dh/da > 0$. An increase in the tax rate *reduces* labor income by wh. The second term must then be the substitution effect because the income and substitution effects sum up to the total effect.

2.3.5 *Valued Government Spending*

The case where government spending is valued is now analyzed. There are two cases to consider. In the first case government spending directly substitutes for private consumption, while in the second case it does not. Once again presume that all tax revenue is used to finance government spending; i.e., $\lambda = 0$.

Government spending is valued in the same way as private consumption

It is easy to allow for government spending to be valued. For instance, one could write the consumer/worker's utility function as

$$U(c+\omega g)+V(1-h),$$

where ω is a constant specifying the value that an agent derives from government spending. Equation (2.3.3) now appears as

$$U_1\Big([1-(1-\omega)\tau]wh+a\Big)(1-\tau)w = V_1(1-h),$$

because $\omega g = \omega\tau wh$ so that $-(1-\omega)g = -(1-\omega)\tau wh$. What happens if $\omega = 0$ or $\omega = 1$? It is easy to see that when $\omega = 0$ the above first-order condition reduces to (2.3.5), where there is an income effect connected with taxation. Alternatively, when $\omega = 1$ it is the same as (2.3.4), so that there is no income effect associated with taxation. In general, when $0 < \omega < 1$ the size of the income effect associated with taxation depends upon how valuable or substitutable government spending is in terms of private consumption. This is governed by the size of ω. The drain on private spending due to taxation is offset by the portion of government spending, $\omega\tau wh$, that is substitutable for private spending. When government spending on goods and services is not very substitutable in terms of private consumption the consumer/worker will feel a bigger loss in terms of private consumption than when it is substitutable.

Government spending is valued in a different way than private consumption

Suppose alternatively that government spending is valued according to the concave utility function $G(g)$ so that the individual's utility function can be written as

$$U(c) + G(g) + V(1-h).$$

Equation (2.3.5) now reads

$$U_1\Big((1-\tau)wh + a\Big)(1-\tau)w = V_1(1-h).$$

Thus, surprisingly, this case can be analyzed in the same fashion as the situation where government spending is a deadweight loss! This makes clear that the income effect from government spending derives from the fact that it reduces private consumption. The loss depends on the degree to which government spending is substitutable for private spending.

2.3.6 Progressive Income Taxation

Let the consumer/worker face a progressive tax schedule where his income tax is given by

$$T(wh),$$

with $T(0) = 0, T_1, T_{11} > 0$. The tax schedule is displayed in Figure 2.3.2. It is convex due to the assumption that $T_{11} > 0$. This implies that taxation is progressive since the tax rate that the person pays on the last dollar earned, or $T_1(wh)$, increases with labor income, wh. The consumer/workers choice problem for h now formulates as

$$\max_h \Big\{ U\Big(wh - T(wh) + a + \lambda\Big) + V(1-h) \Big\}.$$

It is straightforward to calculate that his first-order condition will now read

$$U_1\Big(\underbrace{wh - T(wh) + a + \lambda}_{c}\Big)[1 - \underbrace{T_1(wh)}_{\text{marginal tax rate}}]w = V_1(1-h).$$

Again, observe that this is one equation in one unknown. Note that the marginal tax rate, $T_1(wh)$, is higher than the average one, $T(wh)/wh$, because the tax function is convex. When considering the disincentive effect of distortional taxation it is important to use the marginal tax rate and not the average one. It is the marginal tax rate that governs the substitution effect, not the average one. The difference between average and marginal tax rates will be touched upon in Chapter 4, which discusses why American work more than Europeans.

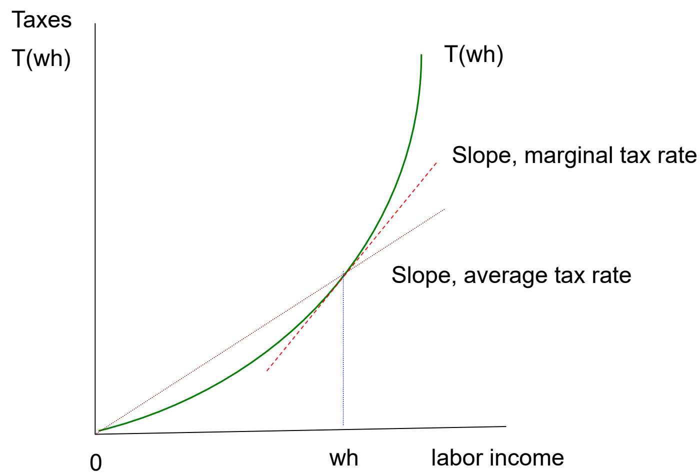


Figure 2.3.2: Relationship between average and marginal tax rates with progressive taxation. The marginal tax rate at the income level wh is given by the slope of the $T(wh)$ curve at the point wh, as shown by the slope of the tangent line. The average rate is illustrated by the slope of straight line from the origin through the point $(wh, T(wh))$, or $[T(wh) - 0]/(wh - 0) = T(wh)/(wh)$. With progressive taxation, the marginal tax rate, $T_1(wh)$, exceeds the average tax rate, $T(wh)/wh$.

2.4 *General Equilibrium*

Production will now be introduced into the above setting with consumers and a government. Once again the government spends g on goods and services, taxes labor income at the rate τ, and provides lump-sum transfer payments in the amount λ. This extension brings into the analysis the notion of a production function. Output, o, will now be produced using capital, k, in addition to labor, h. Capital is assumed to be in fixed supply. This assumption is abandoned in Chapter 6, where the supply of capital is endogenously determined.

Leon Walras (1834-1910) was a French economist at the University of Lausanne. He is best known for his book *Éléments d'économie politique pure* written in 1874. This book founded general equilibrium theory. He derived Walras's law that states the sum of excess demands across markets must sum to zero. This implies that any given market must be in equilibrium, if all other markets are in equilibrium.

2.4.1 *Production Functions*

Assume that output, o, can be produced with capital, k, and labor, h, in line with the following constant-returns-to-scale production function:

$$o = F(k, h).$$

Generally, it is assumed that inputs and outputs are all positive. So, some typical properties imposed on a production function are:

1. $F : \mathcal{R}^2_+ \rightarrow \mathcal{R}_+$ (so that a production function maps the positive reals into the positive reals). Inputs and outputs must always be nonnegative.

2. F is strictly increasing in each of its arguments so $F_1 \equiv \partial F/\partial k > 0$ and $F_2 \equiv \partial F/\partial h > 0$. Marginal products are positive.

3. F is strictly concave in each of arguments requiring $F_{11} \equiv d^2F/dk^2 < 0$ and $F_{22} \equiv d^2F/dh^2 < 0$. This implies that the marginal product for an input decreases with its usage. This is called the law of diminishing marginal returns. Additionally, it must be the case that $F_{11}F_{22} - F_{12}^2 > 0$ so that F is jointly strictly concave in both of its arguments.

4. $F(\theta k, \theta h) = \theta F(k,h)$ for $\theta > 0$–definition of constant returns to scale.[2] This assumption implies that the world can be represented by a single aggregate production function. I.e., it doesn't matter whether the economy has θ firms each producing $F(k,h)$ units of output or one large firm producing $F(\theta k, \theta h)$ units of output. The aggregate levels of inputs and output will be exactly the same.

[2] This property is synonymous with the function being homogenous of degree 1. For the definition of a homogenous function see Chapter A.

Example 2.5. (Common production functions) Here are some production functions that are commonly used in macroeconomics.

$$\begin{aligned}
&F(k,h) = k^\alpha h^{1-\alpha}, \text{ with } 0 \le \alpha \le 1 \text{ (Cobb-Douglas)};\\
&F(k,h) = \alpha k + \beta h, \text{ with } 0 \le \alpha, \beta \text{ (linear)};\\
&F(k,h) = \min\{\chi k, \zeta h\}, \text{ with } 0 \le \chi, \zeta \text{ (Leontief)};\\
&F(k,h) = [\alpha k^\rho + (1-\alpha)h^\rho]^{1/\rho}, \text{ with } 0 \le \alpha \le 1 \text{ and } \rho \le 1 \text{ (constant elasticity of substitution, CES)};\\
&F(k,h) = \alpha k + \beta h - \psi k^2/2 + \delta kh - \varepsilon h^2/2, \text{ with } 0 \le \alpha, \beta, \psi, \delta, \varepsilon, \Delta \equiv \psi\varepsilon - \delta^2 > 0,\\
&\qquad k \le (\alpha\varepsilon + \beta\delta)/\Delta, \text{ and } h \le (\psi\beta + \delta\alpha)/\Delta \text{ (quadratic)}.
\end{aligned}$$

The most commonly used production function in macroeconomics is the Cobb-Douglas one. It will be shown in Chapter 4 that in a competitive equilibrium, α is share of output that is paid to capital and that $1-\alpha$ is labor's share. The linear production is not strictly concave, just concave. With the Leontief production function capital and labor are used in fixed proportions; i.e., $k/h = \zeta/\chi$. For this production function $F_1(k,h) = \chi$, if $k < (\zeta/\chi)h$, and $F_1(k,h) = 0$, otherwise. In other words, when capital exceeds $(\zeta/\chi)h$ extra increments will no longer boost output. A similar property holds for the marginal product of labor. It is easy to check that it satisfies the the constant-returns-to-scale assumption. The CES production function encapsulates the linear, Cobb-Douglas, and Leontief production functions as special cases.

Charles W. Cobb (1875-1949) and Paul H. Douglas (1892-1976) presented the Cobb-Douglas production function in a 1928 article titled "A Theory of Production," which was published in the *American Economic Review*. Cobb had a PhD in mathematics from the University of Michigan and Douglas had a PhD in economics from the Columbia University. Later in life Douglas served as a US senator.

It takes a linear form when $\rho = 1$, converges to Cobb-Douglas form as $\rho \to 0$, and assumes the form $\min\{k, h\}$ as $\rho \to -\infty$.[3] The elasticity of substitution between capital and labor is given by $1/(1-\rho)$. It measures the responsiveness of the capital/labor ratio to a change in the rental/wage ratio. This elasticity is increasing in ρ. Finally, the quadratic production function is not always increasing in k and l so the restriction on the bottom line is needed. It does not satisfy the constant-returns-to-scale assumption, either.

[3] If instead $F(k,h) = [\alpha(\chi k)^\rho + (1-\alpha)(\lambda h)^\rho]^{1/\rho}$, then as $\rho \to -\infty$ the CES production function approaches the Leontief form $F(k,h) = \min\{\chi k, \zeta h\}$.

Kenneth J. Arrow (1921-2017), Hollis B. Chenery (1918-1994), Bagicha S. Minhas (1929-2005), and Robert M. Solow (1924-2023) formulated the CES production in a 1961 paper titled "Capital-Labor Substitution and Economic Efficiency." The paper was published in the *Review of Economics and Statistics*.

2.4.2 *The Firm's Decision Problem*

A firm hires labor to maximize its profits. Let the capital stock, k, be fixed. Denote the rental price of capital by r. A firm's profits are given by it sales, $F(k,h)$, minus what it pays out the factors of production that it hires, $wh + rk$. The firm's profit maximization problem is

$$\max_{h,k}\{\underbrace{F(k,h) - wh - rk}_{\text{profits}}\}.$$

The first-order condition for labor is

$$\underbrace{F_2(k,h)}_{\text{marginal product of labor}} = \underbrace{w}_{\text{marginal cost}}. \tag{2.4.1}$$

The situation is portrayed in Figure 2.4.1. The firm can hire as much labor as it desires at the wage w. The function F_2 implicitly defines the demand for labor. It shows the marginal product of the last unit of labor hired. Due to diminishing returns, or the fact that $F_{22} < 0$, this schedule is downward sloping. The firm hires up to the point where the marginal product of labor equals the wage rate.

The first-order condition for capital has a similar form

$$\underbrace{F_1(k,h)}_{\text{marginal product of capital}} = \underbrace{r}_{\text{marginal cost}}, \tag{2.4.2}$$

where again r is the rental rate for capital.

2.4.3 *Equilibrium*

To characterize the determination of hours worked in the economy (with taxes) solve out for w using (2.4.1) in the consumer/worker's first-order condition for labor (2.3.2) to get

$$U_1\Big((1-\tau)\underbrace{F_2(k,h)}_{w}h + a + \lambda\Big)(1-\tau)\underbrace{F_2(k,h)}_{w} = V_1(1-h).$$

One should solve out for w *after* solving the representative consumer/-worker's problem. Next, substitute out for λ using the government's

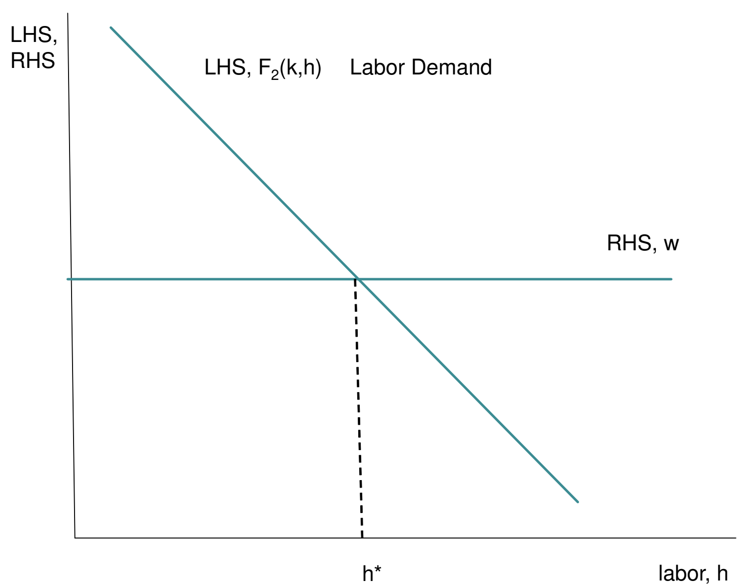


Figure 2.4.1: The firm's hiring decision. The labor demand function is given by the marginal product for labor curve, $F_2(k,h)$. The firm can hire as much labor as it wants at the wage rate w. So, the labor supply function is given by the horizontal line at w. The level of employment, h^*, is determined by the point of intersection between the labor demand and supply curves.

budget constraint (2.3.1) to obtain

$$U_1\Big(F_2(k,h)h + a - g\Big)(1-\tau)F_2(k,h) = V_1(1-h).$$

One might think that the individual owns the economy's fixed capital stock. Each period capital, k, will earn its marginal product, $F_1(k,h)$. Thus, it is reasonable to let $a = rk = F_1(k,h)k$. That is, the worker earns rental on capital accruing from the operation of the firm. This will be discussed in more detail later on. If one solves out for a in this fashion, the result will still be one equation in one unknown:

$$U_1\Big(\underbrace{F_2(k,h)h + F_1(k,h)k - g}_{c}\Big)(1-\tau)F_2(k,h) = V_1(1-h),$$

so that

$$U_1\Big(F(k,h) - g\Big)(1-\tau)F_2(k,h) = V_1(1-h), \qquad (2.4.3)$$

because by the constant-returns-to-scale assumption[4]

$$F_2(k,h)h + F_1(k,h)k = F(k,h) \text{ (Euler's Theorem)}.$$

With constant returns to scale, payments to the factors of production, $F_2(k,h)h + F_1(k,h)k = wh + rk$, completely exhaust output, $F(k,h)$, so that no economics profits are earned; i.e., $F(k,h) - wh - rk = 0$. Now, equation (2.4.3) represents one equation in one unknown, h. By the

[4] See Chapter A for a proof of Euler's theorem.

implicit function theorem, a solution of the form $h = H(k, \tau, g)$ exits. It's easy to see that (2.4.3) fits the form of (2.1.1); just write

$$Z(h) \equiv U_1\Big(F(k,h) - g\Big)(1-\tau)F_2(k,h) - V_1(1-h) = 0. \qquad (2.4.4)$$

So embedded in this single equation are the outcome of the consumer/-worker's consumption-leisure choice problem, the upshot of the firm's profit maximization problem, the government's budget constraint, and market-clearing conditions. Changes in government policy will be reflected by shifts in the form of the function $Z(h)$.[5]

[5] For clarity one could instead write $Z(h; g, \tau)$, which makes clear that the form of the function Z is dependent on g and τ.

2.5 Equivalent and Compensating Variations

To compute the welfare costs of taxation, consider a switch from tax policy regime A to tax policy regime B. Suppose that the representative agent's welfare under policy regime A is given by

$$W^A \equiv U(c^A) + V(1-h^A),$$

where c^A and h^A are his consumption and work effort under this regime. Now, similarly define the person's welfare under policy regime B by

$$W^B \equiv U(c^B) + V(1-h^B).$$

Since utility is an ordinal measure, neither $(W^B - W^A)/W^A$ nor W^B/W^A give a meaningful measure of the welfare gain or loss of moving from policy regime A to policy regime B. To see this, imagine adding a constant term, m, to the representative agent's utility function. This would not affect any of the person's choices. Yet, by making m very large, $(W^B - W^A)/W^A$ can be made arbitrarily small while W^B/W^A would approach one. To get around the ordinal property of utility, Hicks (1941) invented the concepts of compensating and equivalent variations, which measure a person's willingness to pay to make the switch.

2.5.1 Equivalent Variation

Now, how much would a person be either willing to pay or have to be compensated, measured as a fraction of regime A's consumption, to move from A to B? The fraction ϵ solves the equation below

$$U\Big(c^A(1+\epsilon)\Big) + V(1-h^A) = W^B,$$

so that

$$1 + \epsilon = \frac{U^{-1}\Big(W^B - V(1-h^A)\Big)}{c^A},$$

where U^{-1} is the inverse of the function U.[6] The fraction ϵ is called the equivalent variation (EV). Chapter 8 uses the concept of an equivalent variation to compute the welfare cost of business cycles. If $\epsilon > 0$, then the person is willing to pay to move from policy regimes A to B, while $\epsilon < 0$, the person must be compensated.

[6] One can express the first equation as $Z(\epsilon) \equiv U\Big(c^A(1+\epsilon)\Big) + V(1-h^A) - W^B = 0$. So, the problem of finding an equivalent variation amounts to solving one nonlinear equation in one unknown, ϵ.

Example 2.6. (EV with logarithmic Utility) Let $U(c) = \ln(c)$. Then, $U\Big(c(1+\epsilon)\Big) = \ln(1+\epsilon) + \ln c = \ln(1+\epsilon) + U(c)$. For this utility function

$$U\Big(c^A(1+\epsilon)\Big) + V(1-h^A) = \ln(1+\epsilon) + U(c^A) + V(1-h^A) = W^B,$$

implying

$$\ln(1+\epsilon) = W^B - U(c^A) - V(1-h^A) = W^B - W^A,$$

so that

$$\epsilon = \exp(W^B - W^A) - 1.$$

2.5.2 *Excess Burden of Taxation using the Equivalent Variation*

The excess burden of taxation is defined as the welfare cost per unit of extra revenue raised by some proposed form of taxation. Imagine raising the labor income tax by some tiny amount away from tax regime A to a new regime B. The excess burden of taxation is given by

$$\frac{\text{welfare cost of raising taxes}}{\text{change in tax revenue}} = \frac{\epsilon c_A}{\tau_B w_B h_B - \tau_A w_A h_A}.$$

The numerator gives the excess burden of the taxation in terms of consumption units while the denominator gives the amount of new revenue raised (again in consumption units).

Robert E. Lucas, Jr (1937-2023) is one of the most important macroeconomists of the 20th century, along side John Maynard Keynes and Edward C. Prescott. He has written seminal papers in both business cycle theory and endogenous growth theory. Lucas introduced the idea of rational expectations and dynamic general equilibrium modeling into macroeconomics. He also stressed the role of human capital formation for economic growth. In 1995 he was awarded the Nobel Prize in Economics "for having developed and applied the hypothesis of rational expectations, and thereby having transformed macroeconomic analysis and deepened our understanding of economic policy." Like Albert Einstein, Lucas received the prize well after his work was widely recognized as being pathbreaking. Revolutions in ideas don't come easily.

2.5.3 *Compensating Variation*

The notion of a compensating variation (CV) is very similar. Here the question is: How much would a person either be willing to pay or have to be compensated, measured as a fraction of regime B's consumption, to move from A to B? The fraction ψ solves the equation below

$$U\Big(c^B(1+\psi)\Big) + V(1-h^B) = W^A.$$

Example 2.7. (CV with logarithmic Utility) Again let $U(c) = \ln(c)$. Retracing the steps of the previous example while making the appropriate adjustments leads to

$$\ln(1+\psi) = W^A - U(c^B) - V(1-h^B) = W^A - W^B,$$

so that

$$\psi = \exp(W^A - W^B) - 1.$$

Example 2.8. (Importance of economic growth, Lucas (1987, Table 1, p.25)). Consider the impact of changing the rate of growth for an economy. Let a consumer have the lifetime utility function given by

$$W = \sum_{t=1}^{\infty} \beta^{t-1} \ln(c_t), \text{ with } 0 < \beta < 1,$$

where c_t is consumption in year t. Year-t utility, $\ln(c_t)$, is discounted at the rate $0 < \beta^{t-1} < 1$. The further off in the future a utility is, the more it is discounted because β^{t-1} is decreasing in t. The initial level of consumption is c_1. Now suppose that consumption grows at some constant gross rate μ. [7] This implies that $c_2 = \mu c_1$, $c_3 = \mu c_2 = \mu^2 c_1$, $c_4 = \mu c_3 = \mu^3 c_1$, and $c_t = \mu^{t-1} c_1$. Using $c_t = \mu^{t-1} c_1$ in the above lifetime utility function yields

[7] The gross growth rate is one plus the net growth rate. So, if the economy is growing at 3 percent per year, the gross growth rate is 1.03.

$$\begin{aligned} W &= \sum_{t=1}^{\infty} \beta^{t-1} \ln(c_t) = \sum_{t=1}^{\infty} \beta^{t-1} \ln(\mu^{t-1} c_1) \\ &= \sum_{t=1}^{\infty} \beta^{t-1} (t-1) \ln(\mu) + \frac{1}{1-\beta} \ln c_1 \\ &= \frac{\beta}{(1-\beta)^2} \ln(\mu) + \frac{1}{1-\beta} \ln c_1. \end{aligned}$$

Suppose that in regime A consumption grows at the gross rate μ^A while under regime B the growth rate is μ^B. What is the compensating variation associated with a move from A to B, ψ, computed as a fraction of the initial consumption in regime B, c_1^B? A change in the initial level of consumption moves up or down the entire consumption stream. The compensating variation is[8]

$$\psi = \exp[(1-\beta)(W^A - W^B)] - 1 = \exp\{\frac{\beta}{1-\beta}[\ln(\mu^A) - \ln(\mu^B)]\} - 1.$$

[8] The compensating variation solves

$$\begin{aligned} &\underbrace{\beta \ln(\mu^A)/[(1-\beta)^2] + \ln c_1^A/(1-\beta)}_{W^A} \\ &= \underbrace{\beta \ln(\mu^B)/[(1-\beta)^2] + \ln c_1^B/(1-\beta)}_{W^B} \\ &\quad + \ln(1+\psi)/(1-\beta). \end{aligned}$$

Under the assumption that $c_1^A = c_1^B$, the above result obtains.

To make this more concrete, let the annual discount factor, β, be 0.95 and the annual growth rate in regime A, μ^A, be 3 percent. The table below reports the compensating variation for various growth rates in regime B.

SHIFTS IN GROWTH

Growth rate in Regime B, % $(\mu^B - 1) \times 100\%$	CV, % $\psi \times 100\%$
1.0	45
2.0	20
3.0(Regime A)	0
4.0	-17
5.0	-31
6.0	-42

So, a person would be willing to give 17 percent of his regime B's consumption stream to increase growth from 3 to 4 percent–welfare in regime B is higher than in regime A. The person would have have to be given a compensation of 45 percent to move to a situation where the economy grows at only 1 percent. These are large numbers. Lucas's conclusion is that growth effects are important. The welfare cost of business cycles is discussed in Chapter 8. The welfare costs of business cycles turn out to be much smaller, so many economists feel that studying economic growth is more important than studying business cycles.

The pseudo code for the above example is trivial.

Lucas Growth-pseudo code

1. Input values for the discount factor, $\beta = 0.95$, and a vector of gross growth rates, $\vec{\mu} = (1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0, 5.0, 6.0)$.
2. Compute the compensating variation in percentage terms for each value of μ using the formula $\psi = 100 \exp\{[\beta/(1-\beta)][\ln(3.0) - \ln(\mu)]\} - 1$.
3. Display the results in a table format.

2.6 *Solving Nonlinear Equations Numerically*

In *all* of the above labor supply problems the difference between the lefthand and righthand side of the first-order conditions is a decreasing function in h. This occurs because as h increases the left-hand side drops, while the right-hand, which is being subtracted off, moves up. For example, recall equation (2.4.4) which stated

$$Z(h) \equiv U_1\Big(F(k,h) - g\Big)(1-\tau)F_2(k,h) - V_1(1-h). \tag{2.6.1}$$

The function $Z(h)$ is a decreasing function in h because $U_1\big(F(k,h) - g\big)$, $F_2(k,h)$, and $-V_1(1-h)$ are all decreasing in h due to strict concavity assumptions. Again this single equation represents the upshot of a consumer/worker's consumption-leisure choice problem, a firm's profit maximization decision, a government's spending and tax program, and market-clearing conditions. The solution to this model's general equilibrium with taxes and spending then amounts to solving (2.6.1) for the single endogenous variable h. Now assume that $Z(h)$ is a continuous function of a variable h. Suppose that one wants to find numerically the value of h that solves the nonlinear equation

$$Z(h) = 0.$$

This is called finding the zero or root for the function Z. Before proceeding, some properties will be imposed on the function Z, largely for heuristic purposes.

1. $Z : \mathcal{R} \rightarrow \mathcal{R}$. The solution lies in the space of real numbers.
2. $Z(h) = 0$ for some h called h^*. This assumption implies that a solution exists.
3. $Z_1(h) < 0$ for all h. This assumption is imposed for illustration purposes only. The case where $Z(h)$ is an increasing function can be handled by analyzing $-Z(h) = 0$, where $-Z(h)$ will be decreasing in h. The solution for h isn't affected by changing the sign of $Z(h)$.

The above properties imply that

$$Z(h) \text{ is } \begin{cases} > 0, & \text{if } h < h^*; \\ = 0, & \text{if } h = h^*; \\ < 0, & \text{if } h > h^*. \end{cases}$$

The function $Z(h)$ is shown in Figure 2.6.1.

Two methods for numerically solving nonlinear equations are presented here: the bisection algorithm and Newton's method. Pseudo code is presented for each method. Pseudo code follows the conventions of a normal structured programming language. It is intended as a heuristic device, since actual computer code can be difficult to read. Pseudo code does not follow the syntax of any particular programming language.

2.6.1 *Bisection Method*

The bisection method brackets the solution for h, or h^*, between lower and upper bounds, h^l and h^u, so that $h^l < h^* < h^u$. On each iteration of the algorithm one of these bounds is moved toward h^*, which shrinks the bracket. The algorithm is constructed so that h^* always remains within the bounds on each iteration. Eventually h^* is captured within a very tiny bracket, implying that a solution has been found.

The algorithm-pseudo code

Set a desired tolerance for the solution, denoted by $\varepsilon > 0$.

1. Enter iteration j with lower and upper bounds for h^* denoted by $h^{l,j}$ and $h^{u,j}$, such that $Z(h^{l,j}) > 0$ and $Z(h^{u,j}) < 0$.
2. Construct a guess for the solution,

   $$h^j = (h^{l,j} + h^{u,j})/2.$$

   This motivates the name bisection.

3. Check for convergence.

   (a) If

   $$|Z(h^j)| < \varepsilon,$$

   then stop. The desired solution has been found.

   (b) Else

   $$|Z(h^j)| \geq \varepsilon.$$

   Go to Step 4.

4. Construct new lower and upper bounds.

   (a) If

   $$Z(h^j) < 0,$$

   then the solution for h, or h^*, must smaller than h^j, because Z is decreasing in h. Hence, the upper bound, $h^{u,j}$, must be too high. So, set $h^{l,j+1} = h^{l,j}$ and $h^{u,j+1} = h^j$; i.e., reset the upper bound. Return to Step 1.

   (b) If

   $$Z(h^j) > 0,$$

   then set $h^{l,j+1} = h^j$ and $h^{u,j+1} = h^{u,j}$; i.e., reset the lower bound. Return to Step 1.

The process is shown in Figure 2.6.1.

2.6.2 *Newton's Method*

> Nature and nature's laws lay hid in night; God said "Let Newton be" and all was light. (Alexander Pope)

Sir Isaac Newton (1642-1727) is thought by many to be the greatest scientist of all time. He made seminal contributions to astronomy, mathematics, and physics. His method for solving nonlinear equations is the dominant one in numerical analysis even today.

Newton's method has two advantages over the bisection method. First, it is faster, because it uses knowledge about both Z and Z_1 to compute revised guesses. Second, it generalizes easily to systems of nonlinear equations. Its disadvantage is that it can be unstable.

The algorithm-pseudo code

Set a desired tolerance for the solution, denoted by $\varepsilon > 0$.

1. Enter iteration j with a guess for h^* denoted by h^j.

2. Check for convergence

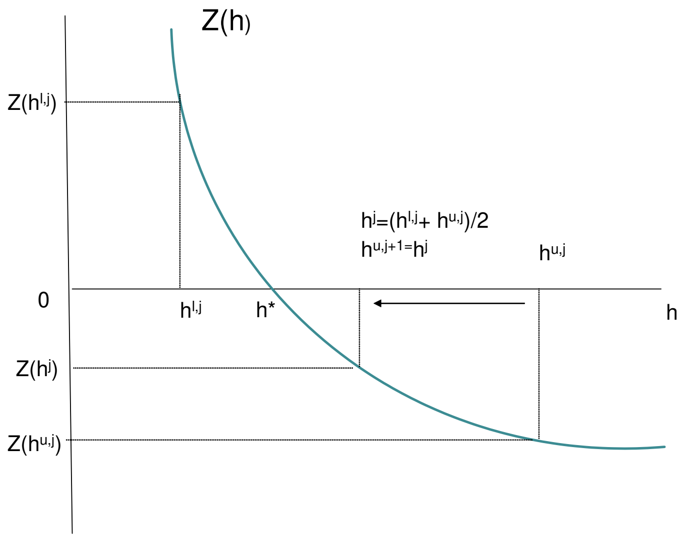


Figure 2.6.1: Illustration of the Bisection Method. Iteration j starts with the lower and upper bounds, $h^{l,j}$ and $h^{u,j}$, on the solution for h. The solution, h^*, is trapped between these bounds, by the Intermediate Value Theorem, because $Z(h^{l,j}) > 0$ and $Z(h^{u,j}) < 0$. (See Chapter A for the Intermediate Value Theorem.) A guess, h^j, is made that bisects this bounds; i.e., that is, $h^j = (h^{l,j} + h^{u,j})/2$. In the situation shown, $Z(h^j) < 0$ so that the true solution must lie below h^j, given the assumption that $Z(h)$ is decreasing. Hence, the upper bound on iteration $j+1$ is lowered by setting $h^{u,j+1} = h^j$.

(a) If

$$|Z(h^j)| < \varepsilon,$$

then stop. A solution has been found.

(b) Else

$$|Z(h^j)| \geq \varepsilon,$$

and go to Step 3.

3. Update guess using

$$h^{j+1} = h^j - \frac{Z(h^j)}{Z_1(h^j)}.$$

Go back to Step 1.

The process is shown in Figure 2.6.2. At the guess h^j one follows the tangent line down to the axis to get the revised guess h^{j+1}. Note that the equation for the tangent line is

$$y = a^j + b^j h^j.$$

So, the revised guess, h^{j+1}, must solve

$$0 = a^j + b^j h^{j+1},$$

implying

$$h^{j+1} = -\frac{a^j}{b^j}.$$

All that is needed is the coefficients a^j and b^j. From the formula for a straight line, $b^j = Z_1(h^j)$. To find a^j note that $Z(h^j) = a^j + b^j h^j$ so that $a^j = Z(h^j) - b^j h^j = Z(h^j) - Z_1(h^j)h^j$. Therefore,

$$h^{j+1} = \frac{\overbrace{-Z(h^j) + Z_1(h^j)h^j}^{-a^j}}{\underbrace{Z_1(h^j)}_{b^j}} = h^j - \frac{Z(h^j)}{Z_1(h^j)}.$$

Newton's method requires knowledge about the derivative $Z_1(h^j)$. Sometimes this can be computed analytically and the formula for the derivative inputted into the nonlinear equation solver. Other times it must be done numerically. Numerical derivatives are discussed in Chapter 9.

Newton's method can be unstable since it is prone to overshooting, as is shown by Figure 2.6.3. Here the algorithm returns a negative revised guess for h for use on iteration $j+1$. This isn't sensible here, as the function Z is not defined when h is negative. Imagine that $Z(h) = U_1(F(k,h) - g)(1-\tau)F_2(k,h) - V_1(1-h)$, as given by (2.4.4). The terms $F(k,h)$ and $F_2(k,h)$ cannot be evaluated at negative values for h. For example, suppose that $F(k,h) = k^\alpha h^{1-\alpha}$ and $F_2(k,h) = (1-\alpha)k^\alpha h^{-\alpha}$. Then a negative value for h would generate complex numbers for these quantities. Newton's algorithm may go awry at this point. This type of problem is often easy to avoid though. To prevent the type of overshooting shown in Figure 2.6.3, sometimes it pays to put a line in the computer code for the function $Z(h)$ stating that $h = \max\{1.0E-8, h\}$. This line binds h above $\underline{h} \equiv 1.0E-8$, and hence zero, and keeps the algorithm from going into the troublesome region. It works, providing that the answer for h is greater than $1.0E-8$. This idea is also shown in the Figure 2.6.3. To bind h between 0 and 1 you could write $h = \min\{\max\{1.0E-8, h\}, 1.0 - .0E-8\}$.[9]

[9] This idea can be implemented in other ways too. For, instance one could let $h = 1/(1+e^x)$ and solve the problem in terms of x. While the variable x can assume a value in interval $(-\infty, \infty)$, the variable h will always lie in the interval $(0,1)$.

2.6.3 *Corner Solutions*

Return to the simple consumption-leisure choice problem poised at the beginning. Suppose that the worker may desire to devote all of his time to the labor force or none of it. That is, perhaps the worker would desire to set $h = 1$ or $h = 0$. Now, the worker's maximization can be expressed as

$$\max_{0 \le h \le 1} \{U(wh + a) + V(1-h)\}.$$

The solution to this problem will have the following form:

$$\begin{aligned} U_1(wh+a)w - V_1(1-h) &= 0, \quad \text{if } 0 < h < 1, \\ U_1(wh+a)w - V_1(1-h) &< 0, \quad \text{if } h = 0, \\ U_1(wh+a)w - V_1(1-h) &> 0, \quad \text{if } h = 1. \end{aligned}$$

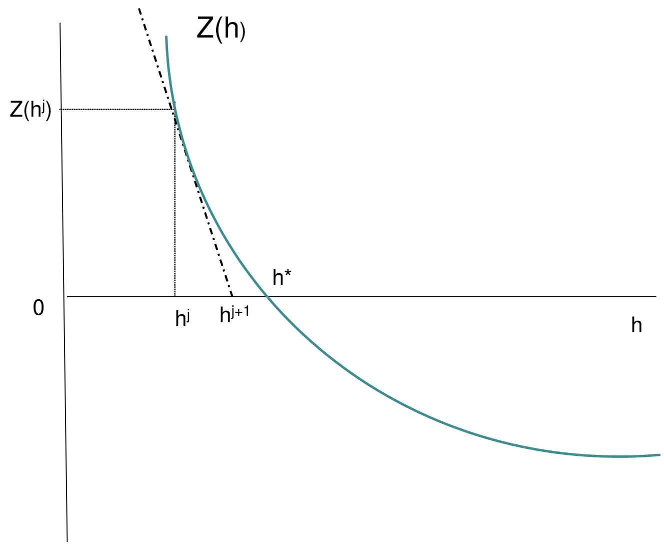


Figure 2.6.2: Illustration of Newton's Method. The guess for the solution on iteration j is h^j. Newton's method uses the line that is tangent to $Z(h)$ at the point h^j to compute the revised guess h^{j+1}. The revised guess is given by the point where the tangent line hits the horizontal axis.

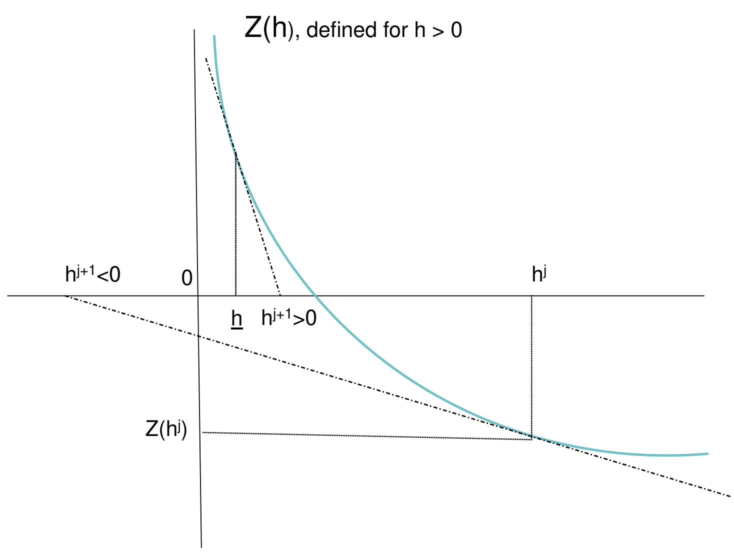


Figure 2.6.3: Example of overshooting when using Newton's method. Here $Z : \mathcal{R}_+ \to \mathcal{R}$, yet on the jth iteration Newton's method returns a negative new guess for h denoted by h^{j+1}. The function Z cannot be evaluated when $h^{j+1} < 0$, since it is not defined for $h < 0$. To prevent overshooting a lower bound, $\underline{h}$, can be placed on the problem. This ensures that for $h^{j+1} > 0$.

The above algorithms can easily be extended to cover this situation. Specifically,

1. Set $h = 0$. Check whether

$$U_1(a)w - V_1(1) < 0.$$

   (a) If so, a solution has been found.

   (b) If not, proceed to Step 2.

2. Set $h = 1$. Check whether

$$U_1(w + a)w - V_1(0) > 0.$$

   (a) If so, a solution has been found.

   (b) If not, proceed to Step 3.

3. Find the zero to the equation

$$U_1(wh + a)w - V_1(1 - h) = 0,$$

   using either the bisection or Newton's method.

Remark 2.2. Let $Z(h) = U_1(wh + a)w - V_1(1 - h)$. One could solve the following equation for h:

$$Z(h)h(1 - h) + h \min\{0, Z(h)\} + (1 - h) \max\{0, Z(h)\} = 0.$$

Note that this equation will return a zero for the true solution.

2.6.4 *Nonlinear Systems of Equations*

Newton's method generalizes easily to systems of nonlinear equations, unlike the bisection method. Consider the nonlinear system of equations

$$\underset{n\times 1}{Z(h)} = \underset{n\times 1}{0}, \tag{2.6.2}$$

where $Z : \mathcal{R}^n \to \mathcal{R}^n$; that is, Z is a system of n nonlinear equations, stacked vertically, in the n unknowns, $h \equiv (h_1, \cdots, h_n)'$. Take a first-order Taylor expansion of the above function around the point h^j, while dropping the remainder term, to get

$$\underset{n\times 1}{Z(h)} = \underset{n\times 1}{Z(h^j)} + \underset{n\times n}{J(h^j)}\underset{n\times 1}{(h - h^j)},$$

where the Jacobian, $J(h^j)$, is a $n \times n$ matrix containing the partial derivatives of Z.[10] The Jacobian is defined by

$$J(h^j) \equiv \begin{bmatrix} Z_1^1(h^j) & \cdots & Z_n^1(h^j) \\ \vdots & & \vdots \\ Z_1^n(h^j) & \cdots & Z_n^n(h^j) \end{bmatrix},$$

[10] This is just a multivariate generalization of the bivariate Taylor expansion reviewed in Chapter A.

where $Z_l^k(h^j)$ refers to the derivative of the k-th row of Z with respect to its l-th argument evaluated at the point h^j. Now, at the h that solves (2.6.2) it transpires that

$$0 = Z(h^j) + J(h^j)(h - h^j),$$

which implies that in a neighborhood around the solution

$$h = h^j - J(h^j)^{-1} Z(h^j).$$

This motivates using the updating equation

$$\underset{n\times 1}{h^{j+1}} = \underset{n\times 1}{h^j} - \underset{n\times n}{J(h^j)^{-1}} \underset{n\times 1}{Z(h^j)}.$$

2.7 A Monopoly Problem

Consider the problem of a monopolist who faces the linear demand function

$$p = \alpha - \frac{\beta}{2} o,$$

where p is the price of the product and o is the monopolist's output. Demand, o, is decreasing in price, p; i.e., $o = (2/\beta)(\alpha - p)$. The monopolist produces according to the quadratic cost function

$$c = \frac{\gamma}{2} o^2,$$

where c is total cost. Marginal cost, γo, is increasing in output, o. In other words, the cost function is strictly convex.

The monopolist's revenue, po, is

$$po = \alpha o - \frac{\beta}{2} o^2.$$

This implies that his profits, π, read

$$\pi = \underbrace{\alpha o - \frac{\beta}{2} o^2}_{\text{revenue}} - \underbrace{\frac{\gamma}{2} o^2}_{\text{costs}}.$$

The monopolist's maximization problem is pick his output to maximize profits. The mathematical transliteration of this maximization problem is

$$\max_o \{\alpha o - \frac{\beta}{2} o^2 - \frac{\gamma}{2} o^2\}.$$

The first-order condition connected with this maximization is[11]

$$\underbrace{\alpha - \beta o}_{\text{MR}} = \underbrace{\gamma o}_{\text{MC}},$$

which sets marginal revenue, MR, equal to marginal cost, MC.

[11] It is easy to calculate that the solutions for output, o, and prices, p, are

$$o = \frac{\alpha}{\gamma + \beta},$$

and

$$p = \alpha - \frac{\beta}{2} \frac{\alpha}{\gamma + \beta},$$

where the solution for o has been substituted into the demand curve to obtain an answer for p.

The numerical problem is to find the zero, o, that solves

$$Z(o) = \alpha - \beta o - \gamma o = 0.$$

Now, set $\alpha = 1.0$, $\beta = 0.5$, and $\gamma = 0.5$. The solution is $o = 1.0, p = 0.75, \pi = 0.5$, and the markup $p/\text{MC} = 1.5$. Figure 2.7.1 shows the solution to the problem. Note that the marginal revenue and marginal cost curves cross at $o^* = 1.0$.

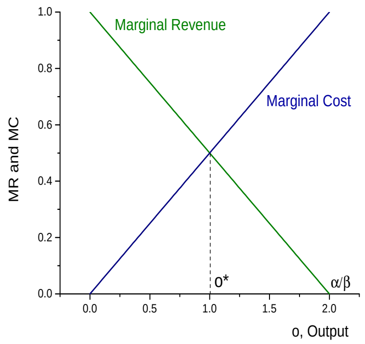


Figure 2.7.1: The marginal revenue and marginal cost curves from a numerical simulation of the monopoly problem. As expected the marginal revenue curve is downward sloping. It hits the x axis at $o = \alpha/\beta$. The marginal cost curve is upward sloping and starts at $o = 0$. The optimal level of output is given by $o^* = 1.0$, which is where the two curves intersect.

Some pseudo code to solve the problem is below.

Monopoly-pseudo code

1. Define α, β, and γ as global variables.
2. Input values for α, β, and γ. Specifically, set $\alpha = 1.0$,$\beta = 0.5$, and $\gamma = 0.5$.
3. Write a function for marginal revenue, $MR(o) = \alpha - \beta o$, taking o as an input and $\alpha - \beta o$ as an output. Pass α and β as global variables into this function.
4. Write a function for marginal cost, $MC(o) = \gamma o$, taking o as an input and γo as an output. Pass γ as global variable into this function.
5. Plot the marginal revenue and marginal cost functions for a grid of output values over the range $[0, \alpha/\beta]$. This gives Figure 2.7.1.
6. Write a function for $Z(o) = MR(o) - MC(o)$ taking o as a variable an returning a value for $Z(o)$. Note that functions can call other functions.
7. Use a nonlinear equation solver to compute a value for o while calling the above function for Z.
8. Check that the function for the first-order condition gives an answer close to zero when evaluated at the computed value for o. That is, check whether $|Z(o^*)|<\varepsilon$, where ε is some desired level of tolerance.

9. Display the results. On this note that that the price is given by $p = \alpha - \beta o^*/2$, the markup by $p/(\gamma o) = [\alpha - \frac{\beta}{2}o]/(\gamma o)$, and profits by $\pi = \alpha o^* - \beta o^{*2}/2 - \gamma o^{*2}/2$.

2.8 *Heterogenous Agents*

Suppose that there are I types of individuals in the economy, namely $i = 1, \cdots, I$. The population of type i agents is of size μ_i, for $i = 1, \cdots, I$. For convenience, set $\sum_{i=1}^{I} \mu_i = 1$. A person has tastes of the following form

$$U(c_i) + V(1 - h_i), \text{ for } i = 1, \cdots, I,$$

where c_i is the consumption enjoyed by a type-i individual and h_i is his work effort. The productivity of person i on the labor market is π_i. Assume that $\pi_I > \cdots > \pi_i > \cdots > \pi_1$, so that person with a higher index has a higher level of productivity. The wage rate for a raw unit of labor is w. A person of type-i will earn the amount $w\pi_i h_i$ in labor income when he works the amount h_i. Suppose that type-i individuals are taxed on their labor income at the rate τ_i. Progressive income taxation implies that $\tau_I > \cdots > \tau_i > \cdots > \tau_1$. There is one unit of capital in the economy. People also earn income from the share of this fixed capital stock that they own. Assume that a type-i agent owns k_i, with $k_I > \cdots > k_i > \cdots > k_1$. Capital earns the rental, r. Production in the economy is undertaken in accordance with the constant-returns-to-scale production function

$$o = F(\mathbf{k}, \mathbf{h}),$$

where $\mathbf{k}$ is the aggregate capital stock and $\mathbf{h}$ is the aggregate input of labor. Last, the government in the economy uses the tax revenue that it collects to distribute lump-sum transfer payments to the populace in the amount λ.

2.8.1 *Type-i's optimization problem*

The optimization problem for a type-i agent is given by

$$\max_{c_i, h_i} [U(c_i) + V(1 - h_i)],$$

subject to

$$c_i = (1 - \tau_i) w \pi_i h_i + r k_i + \lambda.$$

The upshot of this maximization problem is

$$U_1(\underbrace{(1 - \tau_i) w \pi_i h_i + r k_i + \lambda}_{c_i})(1 - \tau_i) w \pi_i = V_1(1 - h_i), \text{for } i = 1, \cdots, I. \tag{2.8.1}$$

2.8.2 The Firm's Problem

The firm's problem is

$$\max_{\mathbf{k},\mathbf{h}}[F(\mathbf{k},\mathbf{h}) - r\mathbf{k} - w\mathbf{h}].$$

This results in

$$F_1(\mathbf{k},\mathbf{h}) = r, \tag{2.8.2}$$

and

$$F_2(\mathbf{k},\mathbf{h}) = w. \tag{2.8.3}$$

2.8.3 General Equilibrium

The government's budget constraint is

$$\begin{aligned}\lambda &= \mu_1\tau_1 w\pi_1 h_1 + \cdots + \mu_i\tau_i w\pi_i h_i + \cdots + \mu_I\tau_I w\pi_I h_I \\ &= \sum_{i=1}^{I}\mu_i\tau_i w\pi_i h_i. \end{aligned}\tag{2.8.4}$$

Since the capital market must clear

$$\mathbf{k} = \sum_{i=1}^{I}\mu_i k_i = 1. \tag{2.8.5}$$

Likewise, market clearing in the labor market gives

$$\mathbf{h} = \sum_{i=1}^{I}\mu_i\pi_i h_i. \tag{2.8.6}$$

To characterize the model's general equilibrium use (2.8.3), (2.8.2) in (2.8.1) to get

$$\begin{aligned}
U_1(\underbrace{(1-\tau_1)F_2(1,\mathbf{h})\pi_1 h_1 + F_1(1,\mathbf{h})k_1 + \lambda}_{c_1})(1-\tau_1)F_2(1,\mathbf{h})\pi_1 &= V_1(1-h_1) \\
&\vdots \\
U_1(\underbrace{(1-\tau_i)F_2(1,\mathbf{h})\pi_i h_i + F_1(1,\mathbf{h})k_i + \lambda}_{c_i})(1-\tau_i)F_2(1,\mathbf{h})\pi_i &= V_1(1-h_i) \\
&\vdots \\
U_1(\underbrace{(1-\tau_I)F_2(1,\mathbf{h})\pi_I h_I + F_1(1,\mathbf{h})k_I + \lambda}_{c_I})(1-\tau_I)F_2(1,\mathbf{h})\pi_I &= V_1(1-h_I).
\end{aligned}$$

Now, note that (2.8.4) and (2.8.6) could be used to solve out for λ and $\mathbf{h}$. The result would be a system of I equations in I unknowns, $h_1, h_2, \cdots, h_I$. Solving this system of equations on the computer may not be an easy business, depending on how large I is. Instead, consider the following algorithm which involves just solving one equation in one unknown at a time.

2.8.4 Algorithm (Walras)

Set a tolerance for the algorithm denoted by ε.

1. Enter iteration j with a guess for the wage rate, w, and transfer payments, λ, denoted by w^j and λ^j. Note that a guess for w amounts to a guess for r because from the equation $w = F_2(1, \mathbf{h})$, so that one can solve for $\mathbf{h}$ and hence r using the relationship $r = F_1(1, \mathbf{h})$.

2. Solve the optimization problems for agents $i = 1, \cdots, I$ using the guesses w^j and λ^j to get a solution for the h_i's. This will involve a FOR or DO loop in the computer program.

3. Calculate what wages and transfer payments are at the solution for the h_i's:
$$w = F_2(1, \underbrace{\sum_{i=1}^{I} \mu_i \pi_i h_i}_{\mathbf{h}}),$$
and
$$\lambda = \sum_{i=1}^{I} \mu_i \tau_i w \pi_i h_i.$$
Compute a revised guess for wages and transfer payment using the formulae
$$(w + w^j)/2,$$
and
$$(\lambda + \lambda^j)/2.$$

4. Check for convergence
   (a) If
   $$|w^{j+1} - w^j|/2 + |\lambda^{j+1} - \lambda^j|/2 < \varepsilon,$$
   then stop.
   (b) Otherwise, return to step 1 with the new guesses.

The topic of heterogenous agents will be returned to in Chapter 10 when the Aiyagari (1994) model is discussed.

2.9 Wassily W. Leontief's Input-Output Table

An input-output describes the flow of goods and services between industries in an economy. Table 2.9.1 shows the basic structure of an input-output table. In the table, industries 1 to m represent the demands for intermediate inputs while industries $m+1$ to n represent the final demands for goods and services. Let $o_{i,j}$ denote the output of industry i that is going to industry j. Moving across a row in the table shows where the output of a particular industry is going to. The sum of outputs across the row denotes the gross output of the industry. Moving down a column gives the intermediate inputs that an industry requires for its production. The sum of these inputs down the column gives the total demand of intermediate inputs required by the industry.

In 1973 Wassily W. Leontief (1906-1999) was awarded the Nobel prize in Economics for his work on input-output analysis. He discusses his work in Leontief (1951).

The input-output table can be used as a simple linear general equilibrium model. The starting point is the creation of an $m \times m$ input/output coefficient matrix, C. To construct this matrix, divide each column in the $m \times m$ matrix for intermediate inputs demand by the total gross output for that sector. Thus,

$$C = \underbrace{\left[c_{i,j} \equiv o_{i,j} / \sum_{k=1}^{n} o_{i,k} \right]}_{m \times m}.$$

Now, the mathematical transliteration of Table 2.9.1 is

$$\underbrace{o}_{\text{gross output}} = \underbrace{Co}_{\text{intermediate inputs}} + \underbrace{f}_{\text{final demand}},$$

where o is a $m \times 1$ vector of gross outputs for each sector with these outputs being used as both intermediate inputs, Co, and as final demands, f, where Co and f are $m \times 1$ vectors. This implies that

$$o = (I - C)^{-1} f \equiv Mf.$$

From this relationship, the production levels, o, of the intermediate goods sectors required to produce the final demands, f, can be computed.[12] The (i, j)th entry in the multiplier matrix $M \equiv (I - C)^{-1}$ speaks to how much sector i's output must be increased in order to produce an extra unit of final demand for sector j and is often called the Leontief inverse. Along the diagonal the coefficients will be bigger than one. To produce an extra unit of final demand for sector j will generally require the use of intermediate inputs from sector j by the other sectors, including j. Figure 2.9.1 displays a portion of Leontief's input-output table for 1947.

[12] The ith rows of the column vectors o and f are $\sum_{j=1}^{n} o_{i,j}$ and $\sum_{j=m+1}^{n} f_{i,j}$.

	Intermediate Demand			Final Demand			Gross Output
Industry	1	$\cdots$	m	m+1	$\cdots$	n	
1	$o_{1,1}$	$\cdots$	$o_{1,m}$	$o_{1,m+1}$	$\cdots$	$o_{1,n}$	$\sum_{j=1}^{n} o_{1,j}$
2	$o_{2,1}$	$\cdots$	$o_{2,m}$	$o_{2,m+1}$	$\cdots$	$o_{2,n}$	$\sum_{j=1}^{n} o_{2,j}$
$\vdots$	$\vdots$		$\vdots$	$\vdots$		$\vdots$	$\vdots$
m	$o_{m,1}$	$\cdots$	$o_{m,m}$	$o_{m,m+1}$		$o_{m,n}$	$\sum_{j=1}^{n} o_{m,j}$
Input Demand	$\sum_{i=1}^{m} o_{i,1}$	$\cdots$	$\sum_{i=1}^{m} o_{i,m}$				
Total Gross Output, $\sum_{i=1}^{m} \sum_{j=1}^{n} o_{i,j}$							

Table 2.9.1: A Generic Input-Output Table

	1 AGRICULTURE AND FISHERIES	2 FOOD AND KINDRED PRODUCTS	3 TEXTILE MILL PRODUCTS	4 APPAREL	5 LUMBER AND WOOD PRODUCTS	6 FURNITURE AND FIXTURES	7 PAPER AND ALLIED PRODUCTS	8 PRINTING AND PUBLISHING	9 CHEMICALS	10 PRODUCTS OF PETROLEUM AND COAL	11 RUBBER PRODUCTS	12 LEATHER AND	13 S
1 AGRICULTURE AND FISHERIES	10.86	15.70	2.16	0.02	0.19	—	0.01	—					
2 FOOD AND KINDRED PRODUCTS	2.38	5.75	0.06	0.01	*	*	0.03	*					
3 TEXTILE MILL PRODUCTS	0.06	*	1.30	3.88	*	0.29	0.04	0.03					
4 APPAREL	0.04	0.20	—	1.96	—	0.01	0.02	—					
5 LUMBER AND WOOD PRODUCTS	0.15	0.10	0.02	*	1.09	0.39	0.27	*					
6 FURNITURE AND FIXTURES	—	—	0.01	—	—	0.01	0.01	—					
7 PAPER AND ALLIED PRODUCTS	*	0.52	0.08	0.02	*	0.02	2.60	1.08					
8 PRINTING AND PUBLISHING	—	0.04	*	—	—	—	—	0.77					
9 CHEMICALS	0.83	1.48	0.80	0.14	0.03	0.06	0.18	0.10					
10 PRODUCTS OF PETROLEUM AND COAL	0.46	0.06	0.03	*	0.07	*	0.06	*					

Figure 2.9.1: The upper left-hand corner of Leonfief's input-output table for 1947. It shows the first 10 entries out of 42 entries. The figures are in billions of 1947\$. The asterisks indicate that the entry was less than \$5 million. Source: Leontief (1951).

As an example take the Bureau of Economic Analysis's input-output table for 1947, reported in 1947\$. This table of 83 industries is reduced to two sectors here. The first sector comprises industries 12 to 64. These are manufacturing industries, with the exception of 12 (maintenance and repair). The second sector includes mining (industries 5 to 10), wholesale and retail (69), and finance and insurance (70). Final demands come from industries 92 to 98, which represent personal consumption expenditure, gross private fixed capital formation, exports, imports, federal government purchases, and state and local government purchases. Table 2.9.2 shows the result. The associated matrix of input-output coefficients is given by Table 2.9.3. The multiplier matrix $M = (I - C)^{-1}$ can be computed using this, as shown in Table 2.9.4.

Now, suppose the federal government increases its demand for Sectors 1 and 2 outputs by 10%. From the BEA's input-output table this will raise the final demand from \$95,000,985 and \$46,219,919 to \$95,310,195 and \$46,230,442, respectively. Using the matrix in Table 2.9.4, this implies that the gross outputs in Sectors 1 and 2 will have to increase from \$176,846,355 and \$65,519,859 to \$177,403,839 and \$65,578,158, respectively. The increase in gross outputs for each sector, \$557,484.1 and \$58,298.7, are bigger than the increase in final demands, \$309,209.6 and \$10,522.7, due to the multiplier effect; i.e., other industries will use some of the output as an intermediate input into their own production.

	Sector 1, Mfg	Sector 2, Non-Mfg	Final Demand
Sector 1, Mfg	77,544,121	4,301,249	95,000,985
Sector 2, Non-Mfg	13,525,810	5,774,130	46,219,919
Gross Output by Sector	176,846,355	65,519,859	

Table 2.9.2: Two-Sector US Input-Output for 1947, in $1947

	Sector 1, Mfg	Sector 2, Non-Mfg
Sector 1, Mfg	0.4385	0.0656
Sector 2, Non-Mfg	0.0765	0.0881

Table 2.9.3: The Two-Sector US Coefficient Matrix for 1947, in $1947

2.9.1 *Leontief's pseudo code*

Define $S1 = \{12, \cdots, 64\}$ to be the set of industries in Sector 1 and $S2 = \{5, \cdots, 10, 69, 70\}$ to be the set of industries in Sector 2. Let $F = \{92, \cdots, 98\}$ represent the set of final demands.

1. Import the 1947 input-output matrix, dubbed `Leontief`, from an EXCEL file.

2. Aggregate to 2×2 input-output matrix.

   (a) Create a function called `Aggregate(source, destination)`, where `source` is a set of source industries and `destination` is a set of industries that uses the output from sources. Feed in the 1947 Leontief table as a global variable. This function aggregates the uses of output from `source`, or the set of i's in `source`, and for each i computes all of the uses by the destination industries, or it sums over all the j's in `destination`. This involves a prototypical nested for loop.

   Prototypical Nested for loop

   ```
   uses = 0
   for i ∈ source
           for j ∈ destination
                   uses = Leontief(i, j) + uses
           end
   end
   ```

   (b) Construct the 2×2 Leontief input-output matrix using the Aggregate function. Represent this matrix by `L2by2`. In particular, `L2by2(i,j) = Aggregate(Si,Sj)`, for $i, j = 1, 2$.

3. Construct a 2×1 aggregated vector of final demands, represented by $f = [f1, f2]'$. Here $f1$ is the sum over all $i \in S1$ of the sums for each i over the columns of final demand for each $j \in F$. So, this is a double sum, which can be computed using a nested for loop.

	Sector 1, Mfg	Sector 2, Non-Mfg
Sector 1, Mfg	1.7985	0.1295
Sector 2, Non-Mfg	0.1509	1.1075

Table 2.9.4: The Two-Sector US Multiplier Matrix for 1947, in $1947

4. Calculate gross output for each sector. To do this, sum over the columns in Leontief to get a 2×1 vector and then add to it f. Call the resulting 2×1 vector of gross outputs o.

5. Compute the input-output coefficient matrix C by dividing all elements in each column j in `L2by2` by the jth component of o.

6. Compute the multiplier matrix $M = (I - C)^{-1}$.

2.10 *Problem*

Consider the problem of a consumer/worker who has tastes of the form

$$\alpha \ln(c + \xi g^{\omega}) + \beta \ln n, \text{ with } 0 < \alpha < 1,$$

where c is his consumption of market produced goods and n is his consumption of home produced goods. The consumer has one unit of time, which he divides between working in the market, h, and at home, $1 - h$. The person earns the amount w on the labor market. The individual also owns one fixed unit of capital that can be rented at the rate r. There is a firm in the economy that produces output, y, according to the production function

$$y = k^{\theta} h^{1-\theta}, \text{ where } 0 < \theta < 1.$$

The firm hires capital and labor from the consumer/worker. Home produced goods are made in line with the household production function

$$n = (1 - h)^{\rho}, \text{ with } 0 < \rho < 1.$$

Last, there is a government in the economy that taxes labor income at the rate τ. It does not tax the income on capital. The government uses the revenue that it raises from labor income taxation to finance government spending, g, and lump-sum transfers payments, λ.

1. Set up and solve the representative agent's and firm's maximization problems. Characterize the economy's general equilibrium.

2. Let $\alpha = 0.4$, $\beta = 0.8571$, $\theta = 0.3$, $\rho = 0.7$, $\xi = 0$, $\omega = 0.6, k = 1.0$ and $\tau = 0.40$. Set $g = 0$. Solve numerically for the worker's level of work effort.

3. Do the following plots as a function of taxation: hours worked, GDP, government revenues (the Laffer curve), utility, and the change in welfare (measured in terms of private consumption). I.e., change taxes by one percentage point intervals over the range, say 0 to 0.99.

4. How do the results change if taxation is used to finance deadweight (non-valued) government spending? (I.e., set $\lambda = 0$ and $\xi = 0$.)

5. Set $\lambda = 0$ and $\xi = 0.9$. What is the economic content of this change in ξ? How big should government be? What are the considerations in determining the optimal size of government?

3 Maximization (and Minimization)

3.1 Introduction

In economics maximization problems are everywhere. The generic maximization problem might take the form

$$\max_h F(h), \tag{3.1.1}$$

subject to

$$G(h) = 0.$$

Here F is an objective function, h is a vector of variables, and G is a function representing a constraint on the choice variables. For example, in a consumer problem F would represent a utility function and G would be the consumer's budget constraint. Note that this problem can be rewritten as

$$\min_h \{-F(h)\},$$

subject to

$$G(h) = 0.$$

So, it is easy to convert a maximization problem into a minimization problem and vice versa.

Often the parameter values for a model are chosen to maximize the model's fit with respect to some data. This is usually done by selecting the parameter values to minimize some objective function containing the model's prediction errors. So, this is another place in macroeconomics where maximization (or equivalently minimization) is important. In macroeconomics this is often done by calibrating a model to match, as close as possible, a set of stylized facts.

Three methods are presented for maximizing functions: golden-section search, discrete maximization, and particle swarm optimization. In the discussion below the constraint on the maximization problems is dropped. There is a wide variety of numerical algorithms, however, that allow for constraints. The discussion then turns to the calibration of economic models. On this two examples are presented: first, the decline in hours worked by males over the last century and

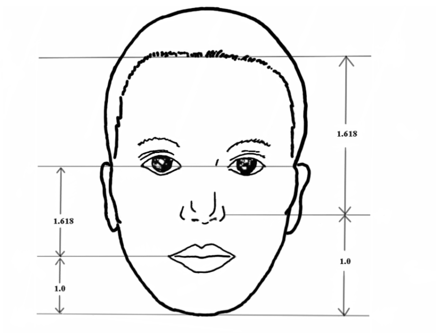


Figure 3.2.1: Beauty and the golden ratio. Is your concept of beauty governed by certain divine facial proportions that are related to the golden ratio? The golden ratio is denoted by the rational number ψ that has the value $1.61803398874\cdots$. In the so-called "ideal" facial structure the ratio of the distance from the eyes to mouth divided by the distance from the mouth to the chin should be ψ. Likewise, ψ should also be the ratio of the distance from the hairline to the bottom of the nose over the distance from the bottom of the nose to the bottom of the chin. George Clooney and Bella Hadid score high when golden ratio formulae are used. Junk science pushed by journalists? Probably.

second the rise in premarital sexual activity by young women over the same period.

3.2 Golden-Section Search

The golden-section search algorithm is reminiscent of the bisection algorithm discussed in Chapter 2. It was invented by the statistician Jack Kiefer in 1953. The algorithm assumes that the objective function $F(h)$ is unimodal–the concept of a unimodal function is defined in Chapter A. This implies that the function F will rise in h until it hits its maximal value and then decline. A unimodal function does not have to be strictly concave because $F_{11}(h)$ does not have to be negative everywhere. Denote the value of h that attains the maximum by h^*. The algorithm starts by imposing a bracket on iteration 1, $[h^{l,1}, h^{u,1}]$, that is known to contain h^* so that $h^* \in [h^{l,1}, h^{u,1}]$. The bracket is successively shrunk until h^* is trapped within some tiny range, at which point the solution has been effectively found, where on each iteration j it transpires that $h^* \in [h^{l,j}, h^{u,j}]$. The rate at which the bracket shrinks is $\alpha = 1/\psi$, where ψ is the golden ratio or the rational number $1.61803398874\cdots$. The golden ratio turns up in architecture, the arts, sciences, and, according to some cosmetic surgeons, perceptions of beauty–see Figure 3.2.1.

3.2.1 The algorithm–pseudo code

The pseudo code for the algorithm is as follows. To start with, set the desired tolerance level for the solution denoted by $\varepsilon > 0$.

1. Enter iteration j with the brackets, $h^{l,j}$ and $h^{u,j}$, around the maximal value such that $h^{l,j} \leq h^* \leq h^{u,j}$.

2. Construct two test points, denoted by $p^{l,j}$ and $p^{u,j}$, which are given by

$$p^{l,j} = h^{l,j} + (1-\alpha)(h^{u,j} - h^{l,j}), \tag{3.2.1}$$

and

$$p^{u,j} = h^{u,j} - (1-\alpha)(h^{u,j} - h^{l,j}) = h^{l,j} + \alpha(h^{u,j} - h^{l,j}). \qquad (3.2.2)$$

where $\alpha = 1/\psi$ and ψ is the golden ratio–see Chapter A. The golden ratio is given by $\psi = (\sqrt{5}+1)/2$ so that $\alpha = 0.61803398874\cdots$. Why this magic number is chosen is discussed below. Note that $p^{l,j} \leq p^{u,j}$, but it does *not* have to be the case that $h^* \in [p^{l,j}, p^{u,j}]$.

3. Update the brackets.

   (a) If
   $$F(p^{l,j}) > F(p^{u,j}).$$
   Since the function is unimodal this implies that $h^* < p^{u,j}$. So the upper bracket, $h^{u,j+1}$, should be reset on the next iteration. In particular,
   $$h^{l,j+1} = h^{l,j} \text{ and } h^{u,j+1} = p^{u,j}. \qquad (3.2.3)$$

   (b) Else
   $$F(p^{l,j}) < F(p^{u,j}),$$
   so that instead the lower bracket is reset on the next iteration, implying
   $$h^{l,j+1} = p^{l,j} \text{ and } h^{u,j+1} = h^{u,j}.$$

4. Check for convergence or that
   $$|h^{u,j+1} - h^{l,j+1}| < \varepsilon.$$
   If the answer is yes, then the solution, h^*, is trapped inside this narrow interval. The algorithm then stops. If the answer is no, return to Step 1.

The golden-section search algorithm is illustrated in Figure 3.2.2, for the case in point 3(a). The test point $p^{l,j}$ is determined so that ratio of the distance from $p^{l,j}$ to it's nearest bound, here $h^{l,j}$, to the distance from $p^{l,j}$ to its farthest bound, $h^{u,j}$, is α. That is, the ratio of a's length to b's length is α. (Or equivalently the ratio of b to a is $1/\alpha = \psi$, the Golden ratio.) The same is true if one instead takes the ratio of the distance from $p^{u,j}$ to its nearest bound to distance from $p^{u,j}$ to its farthest bound. Now, since $F(p^{l,j}) > F(p^{u,j})$ the algorithm resets the upper bound so that $h^{u,j+1} = p^{u,j}$ with the new test point being $p^{u,j+1} = p^{l,j}$. By design this keeps the ratio of the distance of $p^{u,j+1}$ to its nearest bound, now $h^{u,j+1}$, to the distance of the ratio of $p^{u,j+1}$ to its farthest bound, now $h^{l,j+1} = h^{l,j}$, at α. In other words the ratio of c's length to a's length is α.

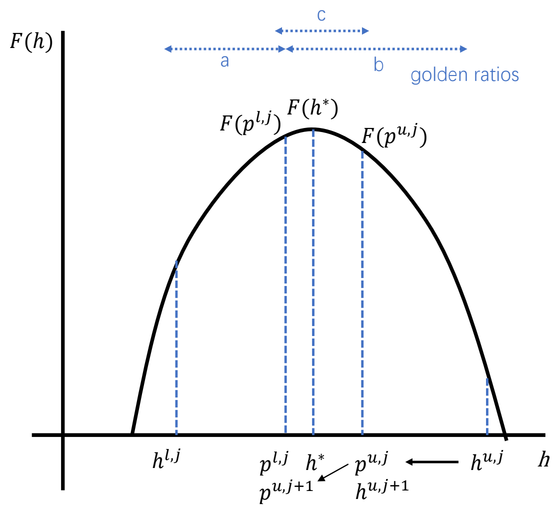


Figure 3.2.2: The golden-section search algorithm. The algorithm starts off in iteration j with the optimal point, h^*, being bracketed in the interval $[h^{l,j}, h^{u,j}]$. Two test points, $p^{l,j}$ and $p^{u,j}$, are then choosen. Because $F(p^{l,j}) > F(p^{u,j})$ it is clear that h^* cannot lie to the right of $p^{u,j}$, because F is unimodal. Hence, on the next iteration, $j+1$, the upper bound can be reset to $h^{u,j+1} = p^{u,j}$. The algorithm is designed so that in iteration $j+1$ the new test point for the upper bound is equal to the old test point for the lower bound; i.e., $p^{u,j+1} = p^{l,j}$. This restriction implies that the line segment ratios, a/b and c/a, are both (the inverses of) golden-section ratios and hence equal to each other. Note that in the situation portrayed in the diagram, $h^* \notin [p^{l,j+1}, p^{u,j+1}]$ because $p^{u,j+1} < h^*$. Unlike as shown, a unimodel function does not have to be strictly concave.

The Determination of α

Exactly how is the constant α determined to achieve this? Observe that on iteration $j+1$ the bracket will be

$$(h^{l,j+1}, h^{u,j+1}) = \begin{cases} (h^{l,j}, p^{u,j}), & \text{if } F(p^{l,j}) > F(p^{u,j}); \\ (p^{l,j}, h^{u,j}), & \text{if } F(p^{l,j}) < F(p^{u,j}). \end{cases}$$

Take the case in point 3(a) where the new bracket for iteration $j+1$ is $(h^{l,j+1}, h^{u,j+1}) = (h^{l,j}, p^{u,j})$; i.e., the upper bound is being adjusted–the second case can be analyzed in a similar manner. Now, suppose the restriction below is imposed where

$$p^{u,j+1} = p^{l,j}; \tag{3.2.4}$$

i.e., the old lower test point becomes the new upper test point. As will be seen, this condition implies that ratio of the distance between the test point and its closest bound to the distance between the test point and its farthest bound is kept constant across iterations. Then, it must happen that

$$\begin{aligned} p^{l,j} &= p^{u,j+1} \\ &= h^{l,j+1} + \alpha(h^{u,j+1} - h^{l,j+1}) \text{ [by updating (3.2.2)]} \\ &= h^{l,j} + \alpha(p^{u,j} - h^{l,j}) \text{ [by (3.2.3)]}. \end{aligned} \tag{3.2.5}$$

Next, using (3.2.1) and (3.2.2) to substitute out for $p^{l,j}$ and $p^{u,j}$ on the left and right, this can be rewritten as

$$\underbrace{h^{l,j} + (1-\alpha)(h^{u,j} - h^{l,j})}_{=p^{l,j}} = h^{l,j} + \alpha[\underbrace{h^{l,j} + \alpha(h^{u,j} - h^{l,j})}_{=p^{u,j}} - h^{l,j}],$$

which by cancelling out the $(h^{u,j} - h^{l,j})$'s reduces to

$$\alpha^2 + \alpha - 1 = 0. \tag{3.2.6}$$

The positive root of this quadratic is $0.61803398874\cdots$, which is $1/\psi$, where ψ is the golden ratio–again, see the Chapter A for more detail.

Observe using (3.2.1) and (3.2.2) that

$$\begin{aligned}\underbrace{\frac{p^{l,j} - h^{l,j}}{h^{u,j} - p^{l,j}}}_{(1-\alpha)/\alpha=\alpha} &= \underbrace{\frac{p^{u,j} - p^{l,j}}{p^{l,j} - h^{l,j}}}_{(2\alpha-1)/(1-\alpha)=\alpha} \\ &= \frac{h^{u,j+1} - p^{u,j+1}}{p^{u,j+1} - h^{l,j+1}} \text{ [using (3.2.3) and (3.2.4)].}\end{aligned}$$

The formula implies that the ratio of the length of the line segment from a bracket to the nearest test point over the length of the other bracket to the same test point is preserved across iterations. So certain ratios are held constant in line with the golden ratio. The above formula motivates the choice of $p^{u,j+1} = p^{l,j}$. To derive the first line in the formula, note that the numerator on the lefthand side is $p^{l,j} - h^{l,j} = (1-\alpha)(h^{u,j} - h^{l,j})$ by (3.2.1). Turn to the denominator on the left. By subtracting (3.2.1) from (3.2.2) it can be seen that $h^{u,j} - p^{l,j} = p^{u,j} - h^{l,j}$. But, (3.2.2) states that $p^{u,j} - h^{l,j} = \alpha(h^{u,j} - h^{l,j})$. Therefore, the ratio on the lefthand side is $(1-\alpha)/\alpha$. The formula for the Golden ratio (3.2.6) implies that this is just α. Now turn to righthand side of the top line. By summing (3.2.1) and (3.2.2) it can be seen that $p^{u,j} - p^{l,j} = (2\alpha - 1)(h^{u,j} - h^{l,j})$. Therefore, the ratio on the righthand side is $(2\alpha - 1)/(1-\alpha)$. Again, this is just α from the golden ratio formula (3.2.6). Thus, the lefthand and the righthand hold with equality at the Golden ratio.[1]

[1] This equation implies that $(1-\alpha)/\alpha = (2\alpha-1)/(1-\alpha)$, so as before α must solve $\alpha^2 + \alpha - 1 = 0$.

Speed of Convergence

Also, note that

$$\begin{aligned}h^{u,j+1} - h^{l,j+1} &= p^{u,j} - h^{l,j} \text{ [using (3.2.3)]} \\ &= h^{l,j} + \alpha(h^{u,j} - h^{l,j}) - h^{l,j} \text{ [using (3.2.2)]} \\ &= \alpha(h^{u,j} - h^{l,j}),\end{aligned}$$

so that bracket is shrinking across iterations by a factor of α. All of this is true for the case in point 3(b).

3.3 General Equilibrium, A Slight Return

Return to the general equilibrium labor supply problem, that was cast in Section 2.4 of Chapter 2. Now, the solution for h cannot be simply found in a single shot by solving one equation in one unknown.

3.3.1 Algorithm (Walras)

Set a tolerance for the algorithm denoted by ε. Suppose the true solution to the problem is h^*. Start with a guess for h^* on the first iteration denoted by h^1.

1. Enter iteration j with a guess for the solution for labor supply denoted by h^j. A guess for h amounts to guesses for r, w, and λ using the equations $r^j = F_1(k, h^j)$, $w^j = F_2(k, h^j)$, and $\lambda^j = \tau F_2(k, h^j)h^j - g$.

2. Solve the maximization problem for the representative agent, *taking as given* r^j, w^j, and λ^j. That is, solve the problem

$$\max_h \{U((1-\tau)w^j h + r^j k + \lambda^j) + V(1-h)\}.$$

   Now, for a revised guess set

$$h^{j+1} = (h + h^j)/2.$$

   The reasoning for this is straightforward. Suppose that the guess, h^j, is too high; i.e., $h^j > h^*$. Then, the representative agent will be receiving too much in transfer payments. He will then work less than he would in equilibrium due to the income effect. So, the solution h will be less than the true solution h^*. Therefore, the true solution must lie between h and h^j. One could think about h as being the supply of hour worked and h^j as the demand for them at the conjectured prices. The true solution should be somewhere in between.

3. Check for convergence

   (a) If

$$|h^{j+1} - h^j| < \varepsilon,$$

   then stop.

   (b) Otherwise, return to step 1 with the new guess.

If the algorithm converges, then by construction an equilibrium will prevail. At the prevailing level of prices, r and w, and transfer payments, λ, the representative agent maximizes his utility. In equilibrium he will earn a wage rate of $w = F_2(k, h)$, which is the marginal product of his labor. He will also earn a rental rate on capital of $r = F_1(k, h)$, which is the marginal product of capital. Last, the person's transfer payments are difference between what the government collects in taxes, τwh, and spends on goods and services, g.

3.4 Discrete Maximization

This is the simplest technique, but the least accurate. Here the domain of the objective function is discretized and the range of the function is evaluated on each point in this discrete set. In particular, suppose that h must lie in the discrete set $\mathcal{H}$. The maximization problem (3.1.1) thus appears as

$$\max_{h \in \mathcal{H}} F(h).$$

3.4.1 The algorithm–pseudo code

The operationalization of discrete maximization is easy

1. Discretize the domain of the objective function to get $\mathcal{H} \equiv (h_1, h_2, \cdots, h_n)$.
2. Construct the vector $\mathcal{F} \equiv (F(h_1), F(h_2), \cdots, F(h_n))$, which represents the associated range of the objective function.
3. Pick the largest element in $\mathcal{F}$, or find $F(h_j)$ such that

$$F(h_j) > F(h_i), \text{ for all } i \neq j.$$

This amounts to just searching a list of numbers and finding the maximal value, something computers can do quickly. If the grid of points in $\mathcal{H}$ is fine enough, then h_j should be reasonably close to the solution, h^*, that obtains from maximization problem where h is allowed to vary continuously. The situation is shown in Figure 3.4.1. Discrete maximization can handle constraints fairly easily. For example, lower and upper bounds on h can be imposed by restricting elements in the set $\mathcal{H}$ to lie in within the range imposed by the bounds. Discrete maximization is often used to solve dynamic programming problems and is returned to in Chapter 9.

3.5 Particle Swarm Optimization

Imagine unleashing a group of bots to find a target, here the global maximum of a function. Each bot starts off from a different position; to wit, a different value of the control variable. A bot modifies its search for the maximum based on two principles. First, it changes its current position in a random manner based on its own personal past best, which is the position in its search history that yielded the highest value of objective function. Second, it also modifies its current position

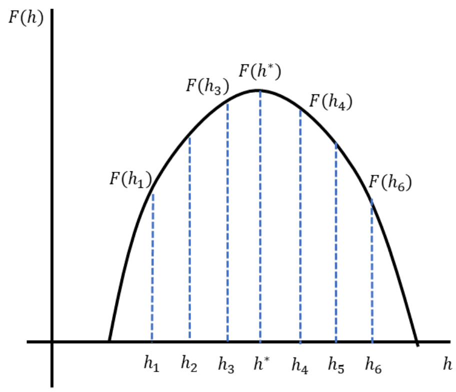


Figure 3.4.1: Discrete Maximization. In the figure the domain for h is converted to a six-point set $\mathcal{H} \equiv (h_1, h_2, \cdots, h_6)$. The function obtains its maximum at $h = h_3$. How close this point is to optimal solution, h^*, depends on the number and spacing of points in the set. By adding more points to set, the solution can be made more accurate. The points can also be made more dense around where the presumed solution lies.

in a random way based upon the best position that all bots have found in their past searches. So, there is both individual and social learning going on. The injection of randomness operates to ensure that various parts of the function are sampled. As the bots continue their searches they will start to swarm toward the global maximum. The algorithm is an application of machine learning. It is said to resemble the behavior of a flock of birds searching and then homing in on a food source. The algorithm was developed by James Kennedy and Russell Eberhart in 1995.

Just two equations are central to the particle swarm algorithm. Suppose that there are I bots. The first equation describes how bot i changes its position between iteration j and $j+1$.

$$h_i^{j+1} = h_i^j + s_i^{j+1}, \text{ for } i = 1, \cdots, I, \tag{3.5.1}$$

where s_i^{j+1} is the step size that the bot will take for iteration $j+1$. The second equation regulates bot i's step size for iteration $j+1$, or s_i^{j+1}, and reads

$$s_i^{j+1} = \underbrace{\alpha \times s_i^j}_{\equiv b} + \underbrace{\beta \times \zeta_i^j \times (h_i^{j*} - h_i^j)}_{\equiv a} + \underbrace{\gamma \times \xi_i^j \times (\mathbf{h}^{j*} - h_i^j)}_{\equiv c}, \text{ for } i = 1, \cdots, I. \tag{3.5.2}$$

In the above step-size equation, α, β, and γ are just constants terms that are fixed across all bots and iterations. The first term, $\alpha \times s_i^j$, is called

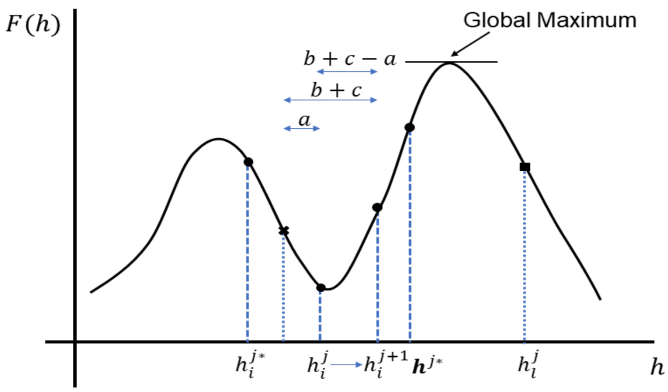


Figure 3.5.1: Particle Swarm Optimization. The original position of bot i on iteration j is shown by h_i^j. The bot moves its position to h_i^{j+1} for iteration $j+1$. The size of the step is $b+c-a$–see equation (3.5.2) for the definitions of a, b, and c. [The a in the diagram is the negative of the one in equation (3.5.2).] The bot's personal best was h_i^{j*}. In accordance with (3.5.2), this induces it to step back by the distance $a = \beta \times \zeta_i^j \times (h_i^{j*} - h_i^j)$. The best position that all of the bots have found is $\mathbf{h}^{j*}$, which is closer to the global maximum. This, in conjuction with the inertial component (the sign of which hasn't been specified), will cause the bot to move forward by $b + c = \alpha \times s_i^j + \gamma \times \xi_i^j \times (\mathbf{h}^{j*} - h_i^j)$. The current position of some other bot, l, is shown by h_l^j.

the inertial component and operates to keep the bot moving in the same direction as taken in previous interation. The second component, $\beta \times \zeta_i^j \times (h_i^{j*} - h_i^j)$, reflects individual learning. Here h_i^{j*} is best position that bot i has personally experienced in the past up to and including iteration j. This term causes the bot to return to the location where it did the best. The coefficient $\zeta_i^j \in [0, 1]$ is a random number that bot i draws on iteration j. This encourages the bot's to search new regions of the space where the control variable lies. The last term, $\gamma \times \xi_i^j \times (\mathbf{h}^{j*} - h_i^j)$, is the social learning component. The variable $\mathbf{h}^{j*}$ is the best position that *any* bot has found in the past. This entices the bot to move to regions where the swarm has found to be productive. Again, there is some randomness in this move since $\xi_i^j \in [0, 1]$ is a randomly drawn number. Figure 3.5.1 portrays the situation.

3.5.1 *The algorithm–pseudo code*

The following steps describe how to operationalize the particle swarm algorithm.

1. Initialize I bots.

   (a) Randomly assign an initial position, h_i^0, in the control space for each of the $i = 1, \cdots, I$ bots.

   (b) Likewise, for each bot i randomly pick an initial step size, s_i^0.

2. On a generic iteration j compute $F(h_i^j)$ using (3.5.1) and (3.5.2) for all $i = 1, \cdots, I$. For this iteration, h_i^{j-1*}, $\mathbf{h}^{j-1*}$, and s_i^{j-1} are all known from the previous iteration, $j-1$.

   (a) Update personal best. If $F(h_i^j) > F(h_i^{j-1*})$, then set $h_i^{j*} = h_i^j$. Else, $h_i^{j*} = h_i^{j-1*}$.

   (b) Update the swarm's best. To do this, find the bot that is doing the best on iteration j. I.e., find

$$m = \arg\max_i \{F(h_1^j), \cdots, F(h_I^j)\}.$$

   If $F(h_m^j) > F(\mathbf{h}^{j-1*})$, then set $\mathbf{h}^{j*} = h_m^j$. Else, $\mathbf{h}^{j*} = \mathbf{h}^{j-1*}$.

3. Decide whether to stop or move onto iteration $j+1$.

   (a) If $F(\mathbf{h}^{j*})$ hasn't improved for J iterations, then stop.

   (b) Otherwise, update each bot i's position according to equations (3.5.1) and (3.5.2).

The particle swarm algorithm is good for global optimization. Its coding is simple and it is perfect for parallel computing, where the problems for many bots can be solved simultaneously. Its main disadvantage is that during the final stages it is slow to home in on the optimal solution. But, for these stages the algorithm could switch to speedier local maximizers, such as golden-section search.[2]

[2] The parameter α should lie between 0 and 1 for convergence reasons. Normally values between 1and 3 are selected for β and γ.

3.6 The Monopoly Problem, A Slight return

Recall the monopoly problem discussed in Chapter 2. The monopolist's problem is to maximize its profits:

$$\max_o \{\underbrace{\alpha o - \frac{\beta}{2}o^2}_{\text{revenue}} - \underbrace{\frac{\gamma}{2}o^2}_{\text{cost}}\}.$$

3.6.1 Golden-Section Search/Particle Swarm Algorithm

This problem can be solved directly using a maximization (or minimization) routine rather than employing a nonlinear equation solver to find the answer from the problem's first-order condition. As before, set $\alpha = 1.0, \beta = 0.5$, and $\gamma = 0.5$. The solution is $o = 1.0$, $p = 0.75$, $\pi = 0.5$, and $p/\text{MC} = 1.5$. Figure 3.6.1 shows the monopolist's objective function.

Monopoly-pseudo code

Some pseudo code to solve the problem is below.

1. Define α, β, and γ as global variables.
2. Input values for α,β, and γ.
3. Write a function for profits, `Profits`(o), taking o as an input and $\alpha o - \beta o^2/2 - \gamma o^2/2$ as an output. Pass α, β, and γ as global variables into this function. If a minimization routine is used, switch the sign of the objective function to read $-\alpha o + \beta o^2/2 + \gamma o^2/2$.
4. Plot the objective function, `Profits`(o), on a grid for output covering the interval $[0, 2]$. This gives Figure 3.6.1.
5. Use a maximization or minimization routine, such as golden-section search or the particle swarm algorithm, to compute value for o while calling the above function `Profits`.
6. Display the results.

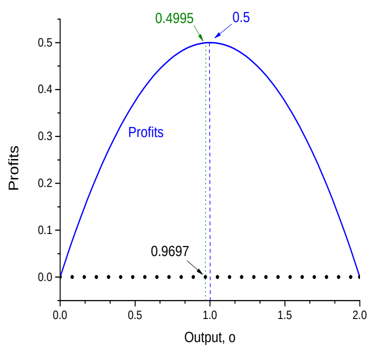


Figure 3.6.1: The monopolist's profits. The optimal level of output is at $o = 1.0$. The figure also shows what happens when a discrete maximization algorithm is employed. The dots represent the grid of output points used in the discrete maximization procedure.

3.6.2 *Discrete Maximization*

Figure 3.6.1 shows what happens when the monopolist's choice set is restricted to be a 25-point evenly spaced output grid between 0 and 1.9394. The point $o = 0.9695$ acheives the maximial value of profits at $0.4995 < 0.5$.

3.7 *Calibration*

Economic theory comes alive when confronted with data. One method of matching economic models with data is calibration. This methodology was introduced into economics by Kydland and Prescott (1982) in a now famous paper. An elementary introduction to calibration is contained in Prescott and Chandler (2008). Calibration often refers to adjusting an instrument, scientific or otherwise, so that it matches some known benchmarks. For example, a guitar can be tuned so that the A string has the Stuttgart pitch of 440Hz. After the guitar has been tuned (calibrated) it can be used to play songs in key with others. In economics the model is treated as an instrument and its parameters can be adjusted so that it matches certain features in the data. After an economic model has been calibrated (tuned) it can be used to conduct policy analysis or thought experiments.

Consider the following economic model

$$o = M(p),$$

where o is a $n \times 1$ vector of output and p is a $m \times 1$ vector of parameters. How should p be chosen? Two criteria are used for selecting parameter values. First, the values for some parameters can be assigned from a priori information. On this, appropriate values for some of the parameters may be available in the economics literature. Also, some parameters might have direct counterparts in the data. Denote the vector of parameters that can be assigned from a prior information by $u \subseteq p$.

Second, the remaining parameters values are chosen so that the model matches, as well as possible, a set of data targets. Let the vector of these parameters be represented by $v \subseteq p$. Hence, $p = (u, v)$. The $j \times 1$ vector of data targets is given by d. Typically, data targets are a set of means, correlations, variances. But, they may also include regression coefficients. The model's output vector, o, must include counterparts for each of the j data targets in the vector d. The model's output may include simulated regression coefficients. The parameters v are picked to minimize the predictions error of the model. Therefore, v solves a minimization problem such as

$$\min_v \sum_{i=1}^{j} [d_i - M_i(\underbrace{u, v}_{=p})]^2,$$

where d_i is the ith data target and $M_i(u, v)$ is the model's prediction for this target. Different criteria could be used for the minimization problem, or for the objective function. Sometimes the data targets can be hit exactly.

Often this is done using a nonlinear equation solver instead of a minimization routine, as the example below show. In particular, certain conditions in the model are engineered by choice of parameter values to hold exactly when evaluated at the data targets. When this can be done for the data targets, the model's prediction will exactly match targets so that $d_i = M_i(u, v)$. So, this procedure can also be thought of as solving the above problem. It is perhaps better to express it as finding the parameter vector v that solves a system of first-order conditions combined with other conditions, denoted by`FOCs`, such that

$$\texttt{FOCs}(v; d, u) = 0.$$

The above system of equation takes as given the vector of data targets, d, and the other parameters, u. Generally, the number of first-order cum other conditions needs to be the same as the number of parameters and data targets. Calibration is a very close cousin of econometrics.

3.7.1 *Selecting parameters values by backsolving*

Sometimes parameter values can be obtained by selecting them so that the model's first-order conditions hold exactly at the observed values in the data. As can example, consider the following consumer/worker problem

$$\max_h \{\theta \frac{(wh)^{1-\rho}}{1-\rho} + (1-\theta)\ln(1-h)\}, \text{ with } 0 < \theta < 1 \text{ and } \rho \geq 0.$$

which has the first-order condition

$$\underbrace{\theta(wh)^{-\rho}}_{\text{IE}} \times \underbrace{w}_{\text{SE}} = \frac{1-\theta}{1-h}.$$

Now, in 1900 the average male worked 63 hours a week. This dropped to only 44 hours in 2018–see Figure 3.7.1. Over this time period real wages rose by a factor of 7.7. Is the above model consistent with these facts? There are both income and substitution effects associated with rising real wages–income and substition effects are discussed in Chapter 2. An increase in real wages implies that more consumption, $c = wh$, can purchased for a given level of work effort. The person will work less on account of the income effect, IE, because he would like to use some of the increase in his standard of living to enjoy more leisure. A climb in the real wage implies that the price of leisure has become more expensive. The substitution effect, SE, states that on this account the person will work more. To get hours worked to fall over time the income effect must dominate the substitution effect.

There are two observations to be targeted; viz, hours worked in 1900 and 2018. There are also two parameter values that need to be selected; namely, θ and ρ. If there are 112 non-sleeping hours in a week, then the desired h's are $0.56 = 63/112$ and $0.39 = 44/112$. If the wage rate for 1900 is normalized to one, the wage rate for 2018 is 7.7. Thus, θ and ρ must solve

$$\theta(1.0 \times 0.56)^{-\rho} \times 1.0 - \frac{1-\theta}{1-0.56} = 0, \tag{3.7.1}$$

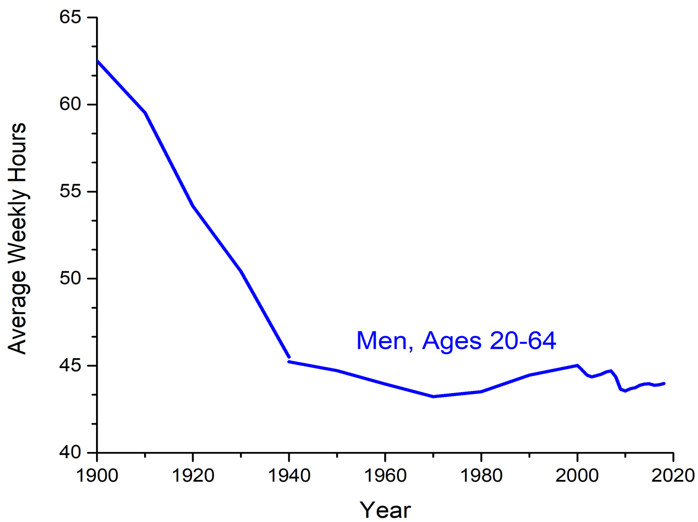


Figure 3.7.1: The plot shows the decline in weekly hours worked by males over the course of the 20th century in the United States. For hours worked to decline the income effect from a rise in real wages must dominate the substitution effect. *Source*: Greenwood et al. (2021b).

and

$$\theta(7.7 \times 0.39)^{-\rho} \times 7.7 - \frac{1-\theta}{1-0.39} = 0. \tag{3.7.2}$$

This represents a system of two equations in two unknowns. Computing the solution using a nonlinear equation solver yields $\theta = 0.50$ and $\rho = 1.41$. These parameter values satisfy the restrictions imposed on them, so that the computed solution is legitimate. Often one can think about the exponent on a function as regulating the change in a variable over time, while the constant term pins down the level in some year given the exponent. To see this, divide the first equation by the second to get

$$[(7.7 \times 0.39)/(1.0 \times 0.56)]^{\rho}/7.7 = (1-0.39)/(1-0.56),$$

so that change in hours worked from 0.56 to 0.39 is governed by ρ, given the change in the wage rate from 1.0 to 7.0. Then, the first equation can be used to solve for θ, given ρ.

Backsolving-pseudo code

1. Define hours-worked data targets, h_{1900} and h_{2018}, and wages, w_{1900} and w_{2018}, as global variables.

2. Input values for hours-worked data targets, $h_{1900} = 0.56$ and $h_{2018} = 0.39$, and values for wages, $w_{1900} = 1.0$ and $w_{2018} = 7.7$.

3. Write a function for the two first-order conditions taking θ and ρ as inputs and the lefthand side of (3.7.1) and (3.7.2) as output. Under the backsolving procedure, the observations h_{1900}, h_{2018}, w_{1900}, and

w_{2018} are parameters. They are passed as global variables into the function for the first-order conditions.

4. Use a nonlinear equation solver to compute values for θ and ρ while calling the above function.

5. Check that the function for the two first-order conditions gives an answer close to zero when evaluated at the computed values for θ and ρ.

6. Check that the values for for θ and ρ are sensible; i.e., whether $0 < \theta < 1$ and $\rho > 0$.

7. Display the results.

3.7.2 *Selecting parameters values by maximizing goodness of fit*

Only 18 percent of 20-year-old women born in 1900 had experienced premarital sex. This rose to 48 percent for women born in 1938 and to 76 percent for those in the 1978 cohort. These facts are shown in Figure 3.7.2. What caused this? Technological progress in contraception is the most likely candidate. The failure rate for contraception in 1900 was 72 percent. This gives the odds of becoming pregnant if a young woman engaged in premarital sex for a year using the available contraception practices at the time. This dropped to 59 percent by 1960 and to 30 percent in 2000.[3]

[3] The failure rates are reported roughly 20 years after the 1938 and 1978 cohorts were born. The number reported for 1900 in Greenwood et al. (2021a) is actually based on data from the 1920's and 30's so it is appropriate to use for the 1900 cohort.

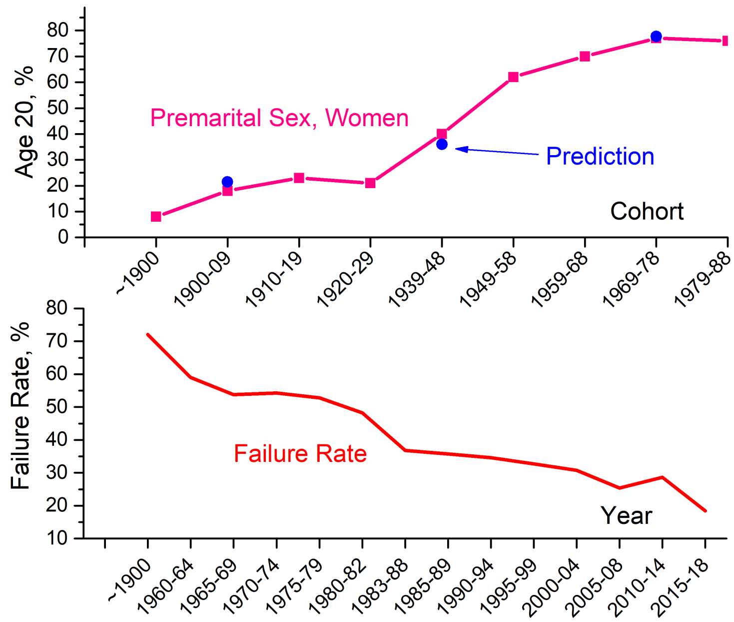


Figure 3.7.2: The chart shows how premarital sexual activity by young women in the United States increased with technological innovation in contraception. *Source*: Greenwood et al. (2021a)

To model this, suppose that the joy a young women gets from a sexual relationship is given by $\tilde{j}$, which is distributed across women according to a Weibull distribution:

$$\Pr[\tilde{j} \leq j] = 1 - \exp[-(j/\eta)^{\beta}], \text{ with } \beta, \eta > 0,$$

where β and η are the shape and scale parameters. (The Weibull distribution is discussed in Chapter A.) Let the cost of an out-of-wedlock birth be represented by O and the failure rate be denoted by f. A young woman's decision to be sexually active is summarized as follows:

$$\begin{aligned} j > fO, &\quad \text{sexually active;} \\ j \leq fO, &\quad \text{abstinent.} \end{aligned}$$

So, a young woman is sexually active when the joy of sex, j, exceeds the expected cost of a pregnancy, fO. The threshold level of joy, j^*, at which a woman is indifferent between having sex or not is given by

$$\underbrace{j^*}_{\text{BENEFIT}} = \underbrace{fO}_{\text{COST}}.$$

At the threshold the joy of sex, j^*, is equal to its expected cost, fO. All women with a level of joy, $\tilde{j}$, above the threshold, j^*, will participate in premarital sexual activity. The fraction of women with premarital sexual experience then reads

$$\Pr[\tilde{j} \geq j^*] = \exp[-(j^*/\eta)^{\beta}] = \exp[-(fO/\eta)^{\beta}].$$

To calibrate this to the U.S. data note that there are 3 parameters, namely β, η, and O. Note that O and η only enter into the minimization problem in the ratio form O/η. Hence, they cannot be separately identified. So, normalize η to be one, as is often done. Observations are at hand for the levels of premarital sexual activity and the failure rates for three years spanning the 20th century. In this case it is impossible to get a perfect fit by solving a system of 3 equations in 2 unknowns. If instead the parameter values are chosen to minimize the model's prediction errors for premarital sex for these three years, then β and O must solve

$$\min_{\beta,O} \Big\{ \{0.18 - \exp[-(0.72 \times O/\eta)^{\beta}]\}^2 + \{0.48 - \exp[-(0.59 \times O/\eta)^{\beta}]\}^2 + \{0.76 - \exp[-(0.30 \times O/\eta)^{\beta}]\}^2 \Big\}. \tag{3.7.3}$$

The solution to this problem yields $\beta = 1.71$ and $O = 2.07$. The circles on the figure show the model's prediction for premarital sex. Since there are two parameters to hit three targets, one would not expect a perfect fit.

Goodness of fit-pseudo code

1. Define premarital sex data targets, p_{1900}, p_{1938}, and p_{1978}, and failure rates, f_{1900}, f_{1938}, and f_{1978}, as global variables.
2. Input values for premarital sex data targets, $p_{1900} = 0.18$, $p_{1938} = 0.48$, and $p_{1978} = 0.76$, and values for the failure rates, $f_{1900} = 0.72$, $f_{1938} = 0.59$, and $f_{1978} = 0.30$.
3. Write a function for the minimand in (3.7.3) taking β and O as inputs and sum of the squares as output. Here the observations p_{1900}, p_{1938}, p_{1978}, f_{1900}, f_{1938},f_{1978}, and η are taken as parameters and passed into the function as global variables.
4. Use a minimization solver to compute values β and O while calling the above function.
5. Report the value obtained for objective function at the solution for β and O. Check that the values for β and O are reasonable.
6. Display the results for the data versus model.

3.7.3 *Hybrid calibration*

Often calibration involves a mixture of these two strategies–an example is Greenwood et al. (2021b). Suppose the vector of fitted parameters, v, is broken down into those solved exactly using first-order and other conditions, x, and those fitted using the minimization routine, y, so that $v = (x, y)$. Let d^x be the vector of data targets used for the former and d^y denote the vector data targets used for the latter. Now the calibration procedure can be represented as

$$\min_y \sum_{i=1} [d_i^y - M_i(\underbrace{u, x, y}_{=p})]^2,$$

subject to

$$\texttt{FOCs}(x; d^x, u, y) = 0.$$

In this procedure the choice of y influences the value of x through the first-order cum other conditions.

4 *Why do Americans Work so Much More than Europeans?*

In general, the art of government consists in taking as much money as possible from one party of the citizens to give to the other. (Voltaire)

4.1 *Introduction*

During the period 1993-96, Americans put in about 50 percent more work than did the French or Italians. Other members of the G7 worked significantly less too. This wasn't always the case. Europeans worked more than Americans over the 1970-74 period. U.S. output per capita is about 40 percent higher than its European counterparts. This is not due to higher productivity, but to higher labor effort. So a question is: Why do Americans work so much more than Europeans?

The answer provided by Prescott (2004) in a classic paper is: European's labor income is taxed at a much higher rate. Prescott calibrates his model to the national income and product accounts for the G7 countries. A short detour through the national income and product accounts is taken since they are an important source of information for macroeconomists. Besides examining the impact of distortional labor income taxation, Prescott's paper raises two other interesting points. First, a consumption tax works in much the same way as a labor income tax does. Second, financing old-age retirement using government mandated private-saving accounts is more efficient than a taxed-financed social security program with lump-sum benefits.

Edward C. Prescott (1940-2022) is the father of quantitative theory. Prescott is the inspiration behind the numerical methods in this book. Before earning his Ph.D. in economics from Carnegie Mellon University, Prescott received a bachelors degree in mathematics from Swathmore and a masters degree in operations research from Case-Western University. With this training he was well suited to bring numerical methods into dynamic stochastic general equilibrium analysis and into economics, more generally. Along with his coauthor and former student Finn E. Kydland, Prescott was granted the Nobel Prize in Economics in 2004. He is one of the foremost macroeconomists of the 20th century.

4.2 *The Model*

To answer this question, let consumer/workers in a country have tastes given by

$$\ln c + \alpha \ln(100 - h).$$

Here each worker is assumed to have 100 hours of non-sleeping time per week.

Output, Labor Supply, and Productivity

output = hours worked × productivity

		Output	*Hours Worked*	*Productivity*
1993–96	Germany	74	75	99
	France	74	68	110
	Italy	57	64	90
	Canada	79	88	89
	United Kingdom	67	88	76
	Japan	78	104	74
	United States	100	100	100
1970–74	Germany	75	105	72
	France	77	105	74
	Italy	53	82	65
	Canada	86	94	91
	United Kingdom	68	110	62
	Japan	62	127	49
	United States	100	100	100

Table 4.1.1: Output, hours worked and productivity in advanced economies.

Output in a country is produced according to

$$o = zk^{\theta}h^{1-\theta}, \tag{4.2.1}$$

where z is a country-specific level of total factor productivity. Take the capital stock in each country to be some fixed number. Suppose that it depreciates at rate δ, with the depreciated portion being made up by gross investment, $i_g = \delta k$. As will be seen, capital doesn't play much of a role in the analysis.

Each country has a government. It spends the amount g on government produced goods and services. It provides transfer payments in the lump-sum amount λ. It taxes labor income at the rate τ_h and consumption at the rate τ_c. The government's budget constraint is

$$\underbrace{\lambda + g}_{\text{expenditure}} = \underbrace{\tau_c c + \tau_h wh}_{\text{revenue}},$$

where w is the wage rate.

Last, there is a resource constraint for a country. It states that

$$c + g + i_g = o,$$

or that consumption, c, plus government spending on goods and services, g, plus gross investment, i_g, equals output, o.

4.2.1 *Worker's Problem*

The representative worker's utility maximization problem is

$$\max_{c,h} [\ln c + \alpha \ln(100 - h)],$$

subject to their budget constraint

$$(1+\tau_c)c = (1-\tau_h)wh + rk + \lambda,$$

where w is the wage rate and r is the rental rate. Observe that the sales tax, τ_c, increases the price of consumption, c.

The first-order condition for labor is

$$\underbrace{\frac{(1-\tau_h)}{(1+\tau_c)}}_{\equiv(1-\tau)}\frac{w}{c} = \alpha\frac{1}{100-h},$$

or

$$\frac{(1-\tau)w}{c} = \alpha\frac{1}{100-h},$$

where

$$\tau \equiv \frac{\tau_c+\tau_h}{1+\tau_c},$$

is the effective tax on labor. The consumption tax, τ_c, creates a disincentive to work, just as the labor income tax, τ_h, does. This makes sense. When deciding how much to work the individual considers the relative price of leisure in terms of consumption goods; i.e., he looks at the forgone consumption that a marginal increase in leisure will cost. Raising the price of consumption, via a consumption tax, reduces the relative price of leisure in a manner similar to increasing the labor income tax.

4.2.2 *Firm's Problem*

The firm's profit maximization problem is

$$\max_{k,h}[zk^{\theta}h^{1-\theta} - rk - wh].$$

Profit maximization for the firm implies

$$w = (1-\theta)zk^{\theta}h^{-\theta},$$

which can be written, using the production function (4.2.1), as

$$w = (1-\theta)o/h.$$

The above equation also implies

$$1-\theta = \frac{wh}{o}.$$

Therefore, $1-\theta$ is labor's share of income.

4.2.3 *Equilibrium–well, almost*

Solving out for the wage rate, w, in the first-order condition for labor yields

$$\frac{(1-\tau)(1-\theta)o/h}{c} = \alpha\frac{1}{100-h},$$

or

$$(100-h)(1-\tau)(1-\theta)o = \alpha ch,$$

so that

$$h = \frac{100(1-\theta)}{\alpha(c/o)/(1-\tau)+(1-\theta)}. \quad (4.2.2)$$

Hours worked, h, is a decreasing function in the effective tax rate, τ, and a decreasing function in the consumption/output ratio, c/o. Loosely speaking the term τ is capturing the substitution effect associated with taxation while the term c/o is tied to the income effect. Anything that increases consumption (relative to output) causes the worker to cut back on his effort since the marginal benefit of working falls. The effect of taxation on c/o will be modest when the revenue from taxation is rebated back as lump-sum transfer payments. In this situation the negative income effect associated with taxation will be minimized.

It is easy to calculate that

$$(1-\tau)\frac{d\ln h}{d\tau} = -\frac{1}{\alpha(c/o)/(1-\tau)+(1-\theta)}\alpha\frac{(c/o)}{1-\tau} < 0.$$

This is the elasticity of labor with respect to a tax change. (Actually, it is the elasticity of labor with respect to $1-\tau$, where it should be noted that $d(1-\tau)/d\tau = -1$). Therefore, the impact of a tax change on labor supply is bigger the larger is c/o and the smaller is $1-\tau$ (or the bigger is τ). The consumption/output ratio, c/o, will be higher when the revenue from taxation is used for lump-sum transfer payments as opposed to wasteful government spending on goods and services. Note that

$$\frac{dh}{d\alpha} = -\frac{100(1-\theta)}{[\alpha(c/o)/(1-\tau)+(1-\theta)]^2}\frac{(c/o)}{1-\tau} < 0. \quad (4.2.3)$$

The higher the weight on leisure, the less people will work, other things equal.

The heart of Prescott (2004) quantitative analysis is equation (4.2.2), which is used to predict hours worked for each country. The following two observations are made about this equation:

1. Given values for the taste and production parameters, α and θ, which will be the same across countries, and observations on the consumption-output ratio, c/o, and the effective tax rate, τ, which will differ across countries, one can make a prediction about hours worked, h, for each country.

2. Strictly speaking the consumption-output ratio, c/o, is an endogenous variable and should ideally be solved out for.

Some examples will now be presented, which echo the theory of labor income taxation presented in Chapter 2. In these examples, let labor be the only factor of production. Therefore, assume a linear production function of the form $o = wh$. Thus, set $\theta = 0$.

Example 4.1. (All government spending is transfer payments) Suppose that the government rebates back tax revenue via lump-sum transfers ($g = 0$). In this case, $c = wh = o$ via the resource constraint (since $g = 0$). Then, the above formula appears as

$$h = \frac{100}{\alpha/(1-\tau)+1}.$$

(Recall that $\theta = 0$.) Hours worked will decline when taxes rise. Only the substitution effect from taxation is present here. There is no negative income effect because all of the tax revenue is rebated back.

(All government spending is a deadweight loss) Alternatively, assume that the government spends all tax revenue. Here, $c = (1-\tau)wh = (1-\tau)o$, using the worker's budget constraint. In this situation the above formula reads

$$h = \frac{100}{\alpha+1}.$$

Taxation has no impact on hours worked because the income and substitution effects exactly cancel out.

(Valued government spending) Now let utility be written as $\ln(c+\xi g) + \ln(100-h)$, with $0 \leq \xi \leq 1$. Assume that the government spends all tax revenue. By parroting the above steps, one now gets

$$(100-h)(1-\tau)o = \alpha(c+\xi g)h.$$

In the above equation use the facts that $c = (1-\tau)wh$ and $g = \tau wh$. This gives

$$(100-h)(1-\tau)o = \alpha[1-(1-\xi)\tau]wh^2.$$

Therefore,

$$h = \frac{100}{\alpha[1-(1-\xi)\tau]/(1-\tau)+1}.$$

If $\xi = 0$, then the result in Example 4.1 obtains, and when $\xi = 1$, the result in Example 4.1 occurs. So, this case is just a hybrid of the previous two cases.

4.3 *National Income and Product Accounts*

The national income and product accounts (NIPA) are a key source of data for macroeconomists. While national income accounting is one of the most boring aspects of macroeconomics, it is also one of the most valuable. NIPA follows the standard practice of double-entry bookkeeping utilized in accounting. On the lefthand side of the accounts is the expenditure on final goods and services produced in the economy. On the righthand side the income earned by consumers, firms, and government. Entries made on the lefthand side must be balanced off with entries on the righthand side, and vice versa. The national income identities for three concepts are explored here; viz, gross domestic product, net domestic product, and national income. The relevant identities are built up piece by piece.

Richard Stone (1913-1991) was instrumental in developing the national income accounts–see Stone and Stone (1966). He introduced the technique of double-entry accounting into the accounts. For all entries on the lefthand (expenditure) side of the ledger there must corresponding entries on the righthand (income) side. For this he was honored with the Nobel prize in 1984. He earned an economics degree from Cambridge University in 1935 and after World War II served as a professor there until his retirement in 1980.

The logic underlying NIPA can be gleaned by thinking about the circular flow of income. Imagine a static setting where final goods are produced just using labor, so income in the economy is just labor income. This labor income is then spent by consumers on final goods. The situation is portrayed in Figure 4.3.1. Let a `CAPITAL` letter denote a variable in the National Income and Product Accounts (NIPA). This gives the NIPA identity where consumption, $\mathtt{C}$, equals labor income, $\mathtt{WL}$:

$$\mathtt{C} = \mathtt{WL}.$$

Now, suppose that there are also profits, Π, on production. Then, the accounts read

$$\mathtt{C} = \mathtt{WL} + \Pi.$$

For some businesses, say sole proprietorships, it is impossible to breakdown income into labor income, $\mathtt{WL}$, and profits, Π. For these types of businesses only proprietors income, $\mathtt{PI}$, is recorded so that now

$$\mathtt{C} = \mathtt{WL} + \Pi + \mathtt{PI}.$$

In reality all income, $\mathtt{WL} + \Pi + \mathtt{PI}$, is not used for domestic consumption, $\mathtt{C}$. Some of it is saved. In particular, people can also use their income for another domestically produced final good, investment, $\mathtt{I}$. There are two concepts for investment, gross and net. Gross investment, $\mathtt{I}_g$, includes spending to replace the depreciation, $\mathtt{D}$, on the existing capital stock. Firms deduct depreciation from their profits, Π. So, if gross investment is used on the lefthand side of the national income identity, then on the righthand side of the national income identity depreciation must be added back because it is deduced when calculating profits yielding

$$\mathtt{C} + \mathtt{I}_g = \mathtt{WL} + \Pi + \mathtt{PI} + \mathtt{D}.$$

Another source of final expenditures is the government, $\mathtt{G}$. They raise the money from taxes, both direct and indirect. Labor income, WL, profits, Π, and proprietors income, $\mathtt{PI}$, are recorded before the direct taxes levied on earned income. Hence, this revenue is already incorporated into the lefthand side of the national income identity.

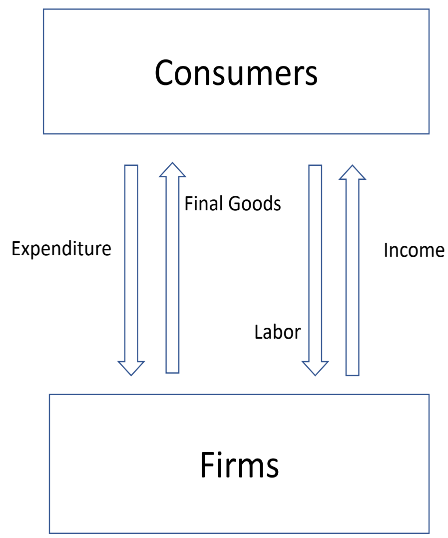


Figure 4.3.1: Circular Flow of Income. Output is produced using solely labor. The income generated from labor is used for expenditure on consumption.

Sales taxes and property taxes, are not. These are called indirect taxes, $\mathtt{IT}$. Indirect taxes are included in the expenditure on the lefthand side and hence they only need to be added to the righthand side to obtain

$$\mathtt{C} + \mathtt{I_g} + \mathtt{G} = \mathtt{WL} + \Pi + \mathtt{PI} + \mathtt{D} + \mathtt{IT}.$$

Additionally, part of expenditure on final goods derives from exports, $\mathtt{E}$. Also, some of expenditure is not on domestically produced goods but on imports, $\mathtt{M}$. Thus, net exports, $\mathtt{E} - \mathtt{M}$, should be added to the lefthand side.

$$\mathtt{GDP} \equiv \mathtt{C} + \mathtt{I_g} + \mathtt{G} + \mathtt{E} - \mathtt{M} = \mathtt{WL} + \Pi + \mathtt{PI} + \mathtt{D} + \mathtt{IT}.$$

The above equation gives the national income identity for *gross* domestic product, $\mathtt{GDP}$. For *net* domestic product, $\mathtt{NDP}$, depreciation is subtracted off of both sides to obtain

$$\mathtt{NDP} \equiv \mathtt{GDP} - \mathtt{D} = \mathtt{C} + \underbrace{I_g - D}_{=\mathtt{I_n}} + \mathtt{G} + \mathtt{E} - \mathtt{M} = \mathtt{WL} + \Pi + \mathtt{PI} + \mathtt{IT},$$

where $\mathtt{I_n} = \mathtt{I_g} - \mathtt{D}$ is net investment. National income, $\mathtt{NI}$, is defined as net domestic product, $\mathtt{NDP}$, less indirect taxes, $\mathtt{IT}$, or

$$\mathtt{NI} \equiv \mathtt{NDP} - \mathtt{IT} = \mathtt{C} + \mathtt{I_n} + \mathtt{G} + \mathtt{E} - \mathtt{M} - \mathtt{IT} = \mathtt{WL} + \Pi + \mathtt{PI}.$$

4.4 Mapping the Model into the Data

Associating model quantities with their analogues in NIPA can be a bit tricky at times. Also, how should parameter values be assigned? The model's resource constraint is $c + g + i_g = o$. Adding indirect taxes on both sides gives $c+\mathtt{IT}+g + i_g = o+\mathtt{IT}$. Consumption and gross investment spending in the data, $\mathtt{C} + \mathtt{I_g}$, include indirect taxes; i.e., $\mathtt{C}= c+\mathtt{IT}_c$ and $\mathtt{I_g}= i_g+ \mathtt{IT_i}$, where $\mathtt{IT_c}$ and $\mathtt{IT_i}$ are the indirect taxes on consumption and gross investment, with $\mathtt{IT_c} + \mathtt{IT_i} = \mathtt{IT}$. Therefore, $\mathtt{C} + \mathtt{G} + \mathtt{I_g} = \mathtt{GDP}$, where $\mathtt{I_g}$ is gross investment. Note that

$$\mathtt{GDP} - \mathtt{IT} = o = zk^{\theta}h^{1-\theta}.$$

..

Assigning Parameter Values

Parameter values will how be assigned for τ_c, τ_h, α, and θ. The notions of consumption and output in the model relative to data are also discussed.

- *Consumption tax,* τ_c. Indirect taxes on consumption, $\mathtt{IT}_c$, are estimated from total indirect taxes, $\mathtt{IT}$, as follows:

$$\mathtt{IT_c} = (2/3 + 1/3 \times \frac{\mathtt{C}}{\mathtt{C} + \mathtt{I_g}})\mathtt{IT}.$$

  The consumption tax rate, τ_c, is then computed as

$$\tau_c = \frac{\mathtt{IT_c}}{\mathtt{C} - \mathtt{IT_c}}.$$

  Note that consumption is approximately $2/3$ of private spending. So, Prescott (2004) is assuming that $2/3$ of indirect taxes fall directly on consumption. Think of this a representing the sales tax on consumption. The remaining indirect taxes fall on business, which will be partially passed on to consumers. This represents things such as gasoline or property taxes, which will be reflected in higher prices for goods and services. So, the remaining $1/3$ is split between consumption and investment according to consumption's share of this spending. Consumption in NIPA is measured as $\mathtt{C} = c + \mathtt{IT}_c$. That is, measured consumption includes the indirect taxes on consumption, $\mathtt{IT}_c$. So, consumption in the model (abstracting from substitutable government spending which is discussed below), c, is given by $c = \mathtt{C} - \mathtt{IT}_c$.

- *Labor income tax,* τ_h. The labor income tax is comprised of two taxes, viz a social security tax, τ_{ss}, and an income tax, τ_{inc}. The social security tax rate is estimated to be

$$\tau_{ss} = \frac{\text{Social Security Taxes}}{(1-\theta)(\texttt{GDP} - \texttt{IT})}.$$

From the the above, $\texttt{GDP} - \texttt{IT} = o$. Therefore, the denominator in the above expression is just labor income. The income tax rate is

$$\tau_{inc} = \frac{\text{Direct Taxes}}{\texttt{NI}}.$$

Direct Taxes includes taxes on interest income. When this measure is used it is impossible to disentangle labor income taxes from taxes on interest income. This explains why total income is in the denominator and not just labor income. Assume that

$$\tau_h = \tau_{ss} + 1.6 \times \tau_{inc}.$$

The number 1.6 translates the average income tax rate into a higher marginal one–recall from Chapter 2 that with progressive taxation the marginal rate must lie above the average one as is illustrated there by Figure 2.3.2. The translation factor is based on the 40% marginal tax rate estimated in Feenberg and Coutts (1993). They take a representative sample of households and see by how much tax revenue will increase if household income rises by 1%. They then estimate the marginal tax rate as (change in tax revenue)/(change in labor income). Assuming that this markup from average to marginal tax rates is valid for all countries is a bit heroic.

- *Mapping consumption and output in the model to the data, c and o.* Effective consumption, c, and output, o, in the model are assumed to have the following relationship with consumption, $\texttt{C}$, and $\texttt{GDP}$ in the data:

$$c = \texttt{C} + \texttt{G} - \texttt{G}_{\texttt{MILITARY}} - \texttt{IT}_{\texttt{c}},$$

$$o = \texttt{GDP} - \texttt{IT}.$$

From the above examples, taxes will have a bigger effect on labor supply the smaller is the income effect. Assuming that nonmilitary government spending, $\texttt{G} - \texttt{G}_{\texttt{MILITARY}}$, is substitutable for private consumption in a one-to-one fashion amplifies the negative impact that taxation has on labor supply.

- *The parameters α and θ.* In standard fashion capital's share of income is set so that $\theta = 0.32$. Labor's share of income, $1-\theta$, is often measured as

$$1 - \theta = \frac{\texttt{WL}}{\texttt{NI} - \texttt{PI}}.$$

How to treat the income from proprietorships is tricky. Some of it will be labor income and some of it will be capital income. The above formula factors out proprietor's income, PI, from national income, NI, to adjust for this. Capital's share of income is assumed to be the same for all countries. The weight on leisure in the utility function, or α, is picked to match average labor supply for the countries. This implies that $\alpha = 1.54$. Prescott (2004) does this in an informal manner, but a more formal fitting procedure, following the discussion on calibration in Chapter 3, yields much the same results, as is discussed below.

4.5 *Actual and Predicted Labor Supplies*

4.5.1 *Historical Episodes*

Under the 1986 U.S. tax reform the marginal tax rate on the second earner in a household dropped. This led to a 10% increase in labor supply between the 1970s and 1990s–recall example 2.2. Most of this was due to an increase in female labor supply. There was a similar reform in Spain in 1998. Labor supply went up by 12%, with a slight increase in revenues.

Example 4.2. (Joint taxation of married household income) Consider the problem of a married couple. Assume that the husband works a fixed 40 hour week, denoted by $\overline{h}$. He earns the wage w and is taxed at the rate τ^1. Let the wife's working hours, h, be flexible. If the woman works, she earns the wage rate ϕw, where ϕ is the gender gap or the ratio of female to male earning. Suppose that if the woman works, then any family income above the level b will be taxed at the higher marginal rate $\tau^2 > \tau^1$–see Figure 4.5.1. That is, when the wife works the family may be pushed into a higher bracket. The household's budget constraint under joint taxation appears as

$$c = \begin{cases} (1-\tau^1)w\overline{h} + (1-\tau^1)\phi w h + \lambda, & \text{if } w\overline{h} + \phi w h < b; \\ (1-\tau^1)b + (1-\tau^2)(w\overline{h} + \phi w h - b) + \lambda, & \text{if } w\overline{h} + \phi w h \geq b. \end{cases}$$

while if they are taxed separately it reads

$$c = (1-\tau^1)w\overline{h} + (1-\tau^1)\phi w h + \lambda.$$

Clearly, joint taxation creates a larger disincentive effect for the second worker, here the wife, than taxing each person separately.

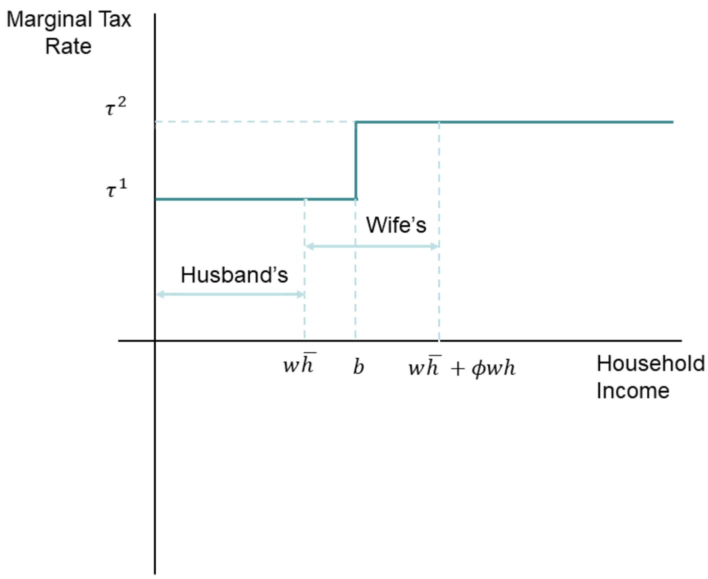


Figure 4.5.1: Joint Taxation. When the wife works the amount h the household is pushed into a higher tax bracket. Any household income above the amount b is taxed at the higher marginal rate $\tau_2 > \tau_1$.

4.5.2 *Results*

The analysis proceeds by using equation (4.2.2) to predict hours worked, h, for each country as a function of their effective tax rate on labor, τ, and their consumption/output ratio, c/o. The two parameters α and θ are common across countries and are chosen in the manner discussed above. Table 4.5.1 below displays the results.

ACTUAL AND PREDICTED LABOR SUPPLY

		Labor Supply		*Differences*	*Prediction Factors*	
		Actual	Predicted		τ	c/o
1993–96	Germany	19.3	19.5	0.2	.59	.74
	France	17.5	19.5	2.0	.59	.74
	Italy	16.5	18.8	2.3	.64	.69
	Canada	22.9	21.3	-1.6	.52	.77
	United Kingdom	22.8	22.8	0	.44	.83
	Japan	27.0	29.0	2.0	.37	.68
	United States	25.9	24.6	-1.3	.40	.81
1970–74	Germany	24.6	24.6	0	.52	.66
	France	24.4	25.4	1.0	.49	.66
	Italy	19.2	28.3	9.1	.41	.66
	Canada	22.2	25.6	3.4	.44	.72
	United Kingdom	25.9	24.0	-1.9	.45	.77
	Japan	29.8	35.8	6.0	.25	.60
	United States	23.5	26.4	2.9	.40	.74

Table 4.5.1: Labor supply.

The model predicts hours worked in the G7 well for the period 1993-96, and less so for the period 1970-74. This can seen better from Figure 4.5.2. The closer a point is to the 45^0 line, the better the fit of the model is with respect to the data. If France reduced its effective tax rate from 60 percent to the U.S. level of 40 percent, it would increase its welfare by 19 percent, measured in terms of consumption. Leisure drops from 81.2 hours to 75.8 hours. There is no reduction in tax revenue. If the United States reduced its income tax rate down from 40 percent to 30 percent welfare would rise by 7 percent in terms of consumption.

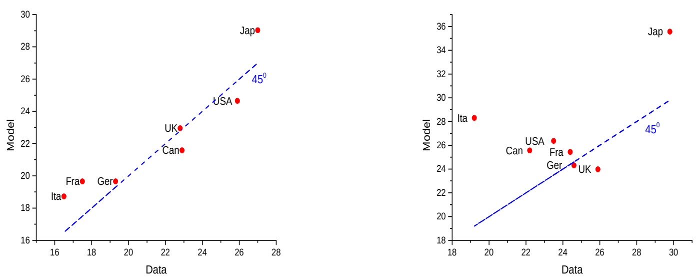


Figure 4.5.2: Sometimes a picture is worth a thousands words. The diagram shows the goodness of fit using Prescott's $\alpha = 1.54$, 1993-96 (left) and 1970-74 (right). If a point lies on the 45^0, then the goodness of fit for that observation is perfect. The further away the point is, the worse the fit. While the model appears to work well for the 1993-96 period, it performs less well for 1970-74.

The pseudo code to solve Prescott's model is straightforward.

Prescott-pseudo code

1. Input data in for the each country's consumption-output ratio, c/o, effective tax rate, τ, and hours worked, h, for both periods. Define these as global variables.

2. Set values for the weight on consumption and capitals share of income: $\alpha = 1.54$ and $\theta = 0.32$. Define these a global variables.

3. Write a function for hours worked using the formula $h = 100(1 - \theta)/[\alpha(c/o)/(1-\tau) + (1-\theta)]$. The function takes the consumption-output rate and the effective tax rate as inputs and outputs a predicted value for hours worked. The two parameters α and θ are defined as global variables.

4. Get the predicted values for hours worked, h, using the function.

5. Construct a table with the results for each country and both time periods. On a plot show the model's predicted values for hours worked versus the values in the data, as in Figure 4.5.2.

The weight on consumption, α, can also be chosen to maximize the fit of the model, as discussed in Chapter 3, Section 3.7. There are 14 data observations. Therefore, α solves a minimization problem of the form

$$\min_{\alpha} \sum_{i=1}^{14} [d_i - M_i(\alpha)]^2,$$

where d_i is the ith data target for hours worked and $M_i(\alpha)$ is the model's prediction for this target. The fit is maximized when $\alpha =$

1.71. The model's fit is not that much different from Figure 4.5.2. So, Prescott's choice for α is close to that which minimizes the above sum of the squares of the prediction errors. The pseudo code for this is below.

Prescott-pseudo code, minimizing the sum of the squares

1. Input data in for the each country's consumption-output ratio, c/o, effective tax rate, τ, and hours worked, h, for both periods. Define these as global variables.

2. Set the value for capitals share of income, $\theta = 0.32$. Define this as a global variable.

3. Write a function for the sum of the squares. This function takes α as an input are returns the sum of the squares as output. The parameter θ is inputted in as a global variable. The data observations are also imputed in as global variables. This gives the d_i's. The model's predictions for each country and time period are given using the formula $h = 100(1-\theta)/[\alpha(c/o)/(1-\tau)+(1-\theta)]$. This gives the $M_i(\alpha)$'s.

4. Call up a minimization routine to compute α using the function for the sum of the squares.

5. Calculate the predicted values for hours worked, h, using the computed value for α.

6. Construct a table with the results for each country and both time periods. On a plot show the model's predicted values for hours worked versus the values in the data, similar to Figure 4.5.2. The figure doesn't change much.

4.5.3 *Financing Social Security*

Suppose that the United States switches to a system of private accounts for social security. In particular, let each worker *choose* between putting 8.7 percent of income into a private retirement account or staying in the present system. The first option reduces the effective income tax rate for a worker to 31.3 percent from 40 percent. This is because, unlike the current system, a worker realizes that he will get these contributions back when he retires; i.e., his retirement payments will now be directly tied to his own work effort. The welfare gain to a 21 year old is estimated to be worth 4 percent of his lifetime consumption. The benefit would be larger still, if he had allowed the retirement age to adjust. That is, in a private system there is more incentive to work longer because this allows you to build up your retirement account.

Example 4.3. (Private-savings accounts) Imagine a individual who lives for two periods. In the first period he works, while in the second period he is retired. Suppose that both the individual and the government face the common interest rate r. Under the current system, the person faces a social security tax on their labor income at the rate τ_{ss} and will receive a lump-sum benefit (essentially unrelated to the work effort) in the amount λ. The person's budget constraint will appear as

$$c_1 + \frac{c_2}{1+r} = (1-\tau_{ss})wh + \frac{\lambda}{1+r},$$

where c_1 and c_2 are consumption in the first and second periods. Clearly, with this budget constraint, the individual's labor-leisure choice will be distorted by the presence of the tax. Under private accounts the person will recognize that $\lambda = (1+r)\tau_{ss}wh$; i.e., that is he will internalize the fact that he will get back with interest any money that goes into his private account. Substituting out for λ in the above budget constraint then gives

$$c_1 + \frac{c_2}{1+r} = wh.$$

Hence, a system of private accounts does not distort the person's labor-leisure choice.[1]

[1] Think about how this relates to Remark 2.1 in Chapter 2.

4.6 *Conclusions*

The results are sensitive to the implied elasticity of labor supply with respect to taxes. This elasticity depends how the revenue from taxation is used because this governs the relative strength of the income and substitution effects from a shift in taxation. A high elasticity implies that hours worked will be very responsive to changes in taxation. A high elasticity occurs when the revenue from taxation is either rebated back in the form of lump-sum transfer payments or when government spending on goods and services is highly substitutable with private spending. In these two cases the income effect associated with taxation is mitigated so that just the substitution effect remains. The deadweight loss from labor income taxation will be high. A high elasticity of labor supply implies that reforms to the social security system, such as private-saving accounts, will significantly improve welfare. Such accounts encourage people to work longer.

5 Graphing

A picture is worth a thousand words.

5.1 Introduction

Graphs can make a paper or presentation come alive. Tufte (2001) says that Charles Joseph Minard's (1781-1870) graph portraying the fate of Napoleon's army during its invasion of Russia in 1812 "may well be the best statistical graph ever drawn." The graph combines a map with a time-series showing the size of Napoleon's army as it travelled from the Polish-Russian border to Moscow and back. The width of the thick red line illustrates the size of the army as it advances towards Moscow. The width of the black line shows the size of it as it retreats. The brutal temperatures facing the army are shown in the bottom panel, which is in sync with the upper one.

5.2 William Playfair

The father of statistical graphing was William Playfair (1759–1823). He is credited with inventing times-series plots, bar graphs, and pie charts. One of Playfair's time-series plots is displayed in Figure 5.2.1. It shows the trade balance over time between England, on the one hand,

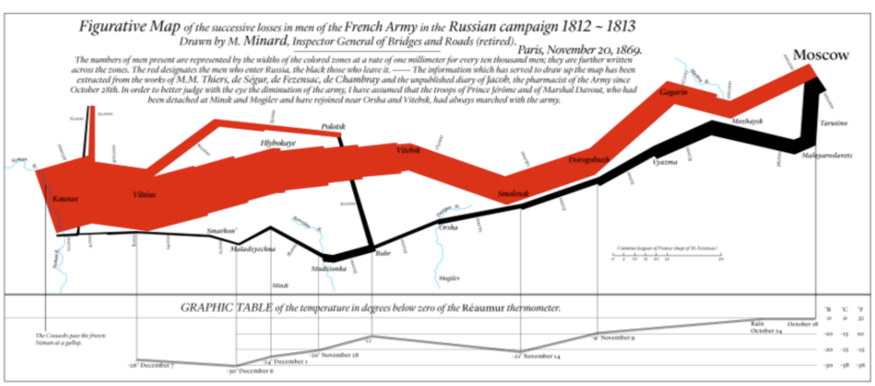


Figure 5.1.1: Napoleon's march. A rendition in english of Charles Joseph Minard's 1869 famous chart.

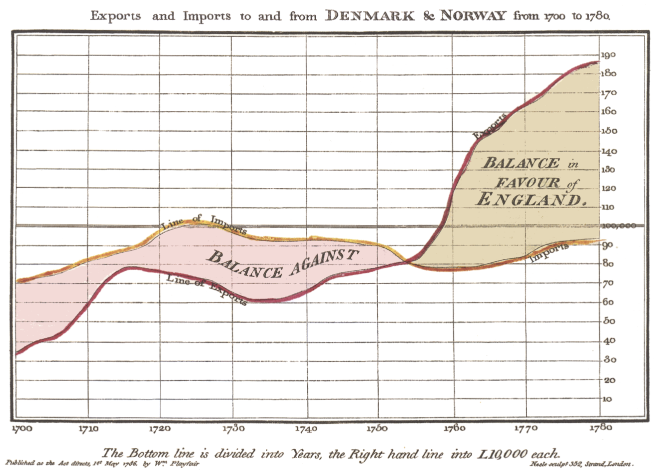


Figure 5.2.1: Source: Playfair, William. *Commercial and Political Atlas*, 1786.

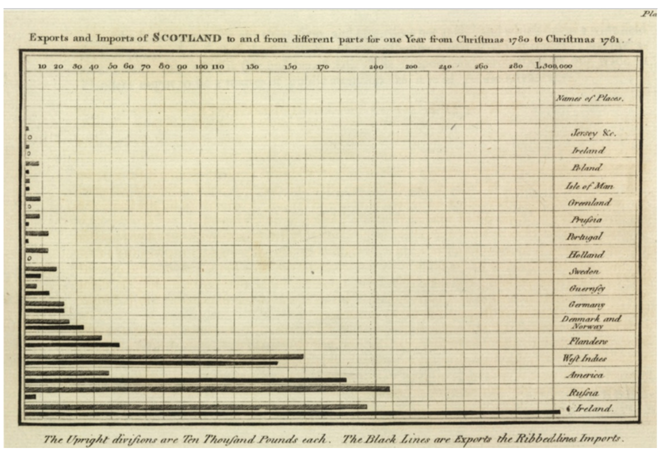


Figure 5.2.2: Source: Playfair, William. *Commercial and Political Atlas*, 1786.

and Denmark and Norway, on the other. Figure 5.2.2 is a bar graph illustrating Scottish exports to and imports from various countries for the year 1781. One of Playfair's pie charts exhibiting the fractions of the Turkish empire (before 1789) located in Africa, Asia, and Europe is presented in Figure 5.2.3.

5.3 Some Basic Principals

Some basic principals for graphing are:

1. *Truthfulness*. Graphs should truthfully display the data. While it's okay to pitch an idea with enthusiasm to an audience or readers, do not change the makeup of a graph to unduly influence people. For example, dips and spikes in time-series plots can be exaggerated by changing the ratio of the vertical to horizontal axes. This is the type of trick that journalists do to flog a story to readers. If

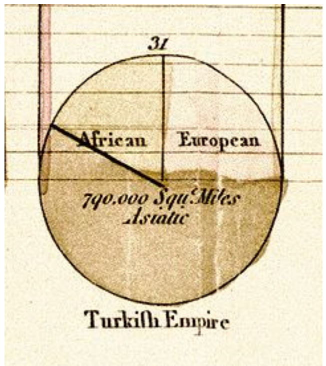


Figure 5.2.3: Source: Playfair, William. *Statistical Brevity*, 1801.

there is choice of data to present, then show what is representative. Of course, your source for the data or graph should be given somewhere. Replication is an important principal in science.

2. *Purposefulness*. A graph should be informative not just a picture. Whatever is being shown should be relevant for the story being told. Graphs should be used to illustrate evidence, an idea or a hypothesis in a paper or presentation, which would be communicated less clearly or forcefully without a graph. As such, they should always be clearly explained in the main text of a paper and or verbally in a presentation.

3. *Clarity*. Graphs should be clear and easy to follow.

   (a) *Captions, colors, fonts, labels, and lines*. Axes should be labeled, in fonts large enough to read, and graphs should have titles in captions. Table 5.3.1 shows how fonts can be use to emphasize something. The caption should also explain the graphical construct, if needed. On a time series plot one can distinguish between the lines by using different line colors and styles. The same is true for the bars on a histogram portraying different data objects. They can be distinguished using different colors and fill patterns. Additionally, make sure that different colors reproduce well in black and white, if this is a requirement.

A Couple of Axes

The mathematician	**T**he ma**t**hema**t**ician
Plotting his past relations	Plo**tt**ing his pas**t** rela**t**ions
"ex" and "why" axis	"ex" and "why" axis

Table 5.3.1: The most commonly used consonant in English is the letter t. By using a bold font the mind can more quickly see this on the version of the haiku shown on the right. This is a take on an example in Schwabish (2014).

(b) *Data-ink maximization*. Tufte (2001) advocates the principal of data-ink maximization. He defines the data-ink ratio as

$$\text{Data-ink ratio} = \frac{\text{data ink}}{\text{total ink used to produce the graph}}.$$

The idea is that most of the ink on a graph should be used to provide information about the data. As such, he recommends deleting boxes around graphs and grids on graphs, since these introduce unnecessary ink and detract from the information shown. Likewise, legends can often be avoided by labeling things directly, such as lines. Does a legend really need a box? Figure 5.3.1 illustrates the idea with two versions of the same of graph. The graphs show the rise in U.S. female-labor force participation in the 20th century from 7 percent in 1860 to 74 percent in 2018. The graph on left uses an antiquated grid, puts the plot in a frame, and includes a legend. The one of the right is much cleaner. Additionally, the plot lines have been thickened to emphasize them and the fonts have enlarged to make them more readable. The range of the x axis has been adjusted to more relevant years.

(c) *Multipanel graphs*. Multipanel graphs containing a large number of subgraphs should be avoided. The subgraphs tend to be small and hard to read.

4. *Aestheticism*. While beauty is in the eye of the beholder, try to make your graph as appealing as possible for the intended audience, without sacrificing the principles of clarity and truthfulness. For example, pleasing coloring schemes can be used. Or colors can be used to represent things such as a national colors or female and male.

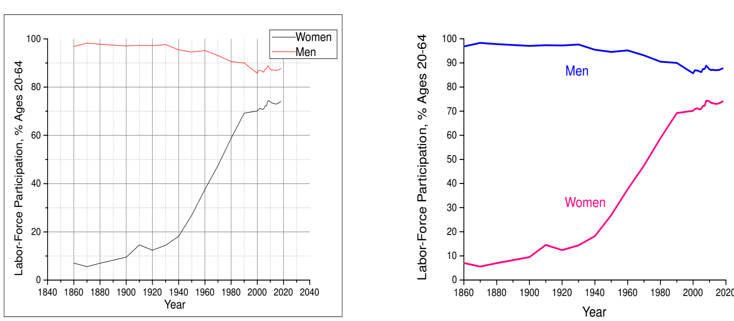


Figure 5.3.1: The graphs show the rise in U.S. female labor-force participation over the 20th century. The left panel presents a graph with a low data-ink ratio. The right panel illustrates the same graph with a high data-ink ratio. *Source*: Greenwood et al. (2021b)

5.4 Modern Renditions of Playfair's Graphs

Modern graphing software makes short work of time series plots, bar charts, and pie charts. What would have taken some time in Playfair's era can be done quickly today.

5.4.1 Time Series Plot

Figure 5.4.1 is a modern version of Playfair's time series export-import plot but applied to United States instead of England.

Time series-pseudo code

1. Import data for exports and imports. This may come in the form of comma separated values (csv) file.
2. Plot the time series data.
3. Make a title, create a legend, and label the x and y axes.
4. Shade the polygon between the import and export curves.
5. Create annotations to label the areas of interest.
6. Save graph in some format; e.g., pdf, png, wmf, etc.

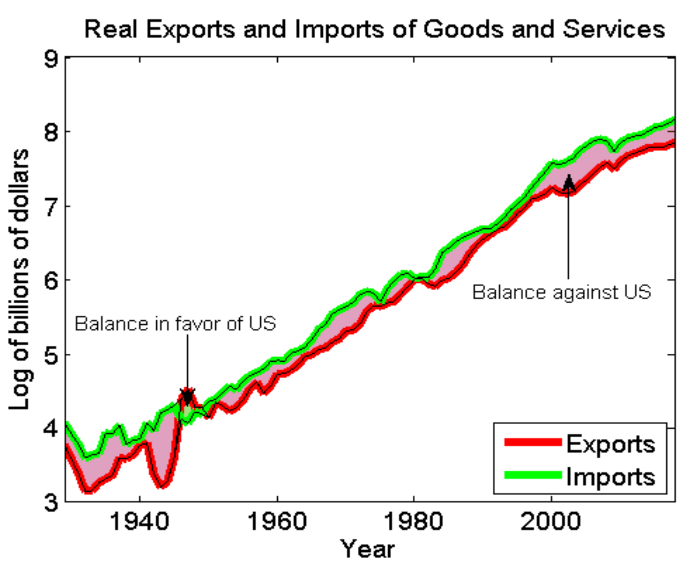


Figure 5.4.1: The version of Playfair's time series plot for the United States that was generated from a MATLAB program. The program was written by Artem Kuriksha.

5.4.2 *Bar Chart*

Playfair's export-import bar chart is generated in Figure 5.4.2 but applied to United States instead of Scotland.

Bar Chart-pseudo code

1. Load in the export and import data for each of the 15 countries. Also load in the names for each country,
2. Plot the data using a grouped horizontal bar chart.
3. Label each country.
4. Make a title and create a legend for the bar chart.
5. Save graph is some format; e.g., pdf, png, wmf, etc.

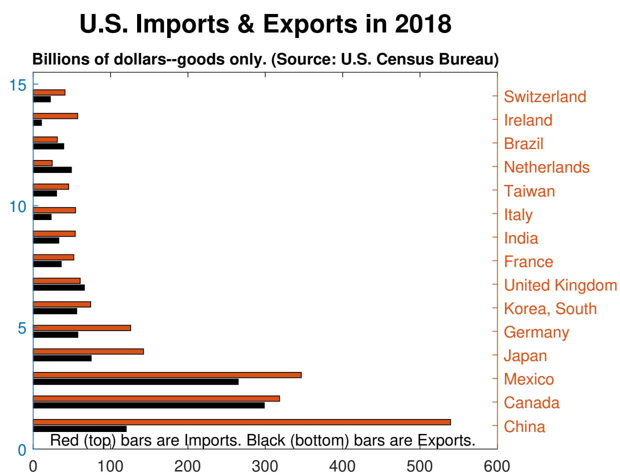


Figure 5.4.2: The version of Playfair's bar chart for U.S. exports and imports in 2018. The MATLAB code was written by Giorgio Lo.

5.4.3 *Pie Chart*

Here is a version of Playfair's pie chart but applied to North America instead of the Turkish Empire–see Figure 5.4.3.

Pie Chart-pseudo code

1. Load in the areas and names for Canada, Mexico, and the United States.
2. Plot the data with the country names using a pie chart.
3. Make a title and a legend.
4. Create annotations to label the three areas of interest.

5. Save graph is some format; e.g., pdf, png, wmf, etc.

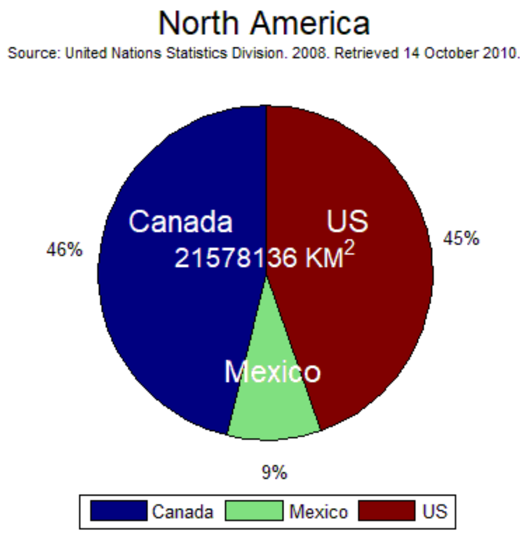


Figure 5.4.3: The version of Playfair's pie chart for North American that was generated from the MATLAB program. The program was written by Giorgio Lo.

6 Deterministic Dynamics

6.1 Introduction

Often macro models specify that the state of economy, k_t, evolves according to a second-order nonlinear difference equation of the following form:

$$k_t = D(k_{t-1}, k_{t-2}), \text{ for } t = 3, 4, \cdots . \tag{6.1.1}$$

The function D may represent the upshot of individuals' and firms' dynamic choice problems, government policies, and market-clearing conditions. By specifying the economy at an elemental level it is hoped the function D will capture the behavioral changes of firms and individuals to the state of the economy, as summerized by the capital stock, technology, and government policy.

To start this difference equation off at time $t = 3$, one would need to know both k_1 and k_2. In period 1 it is reasonable to assume that state of the economy has been predetermined, say at k_{initial}. So, one can employ the starting condition

$$k_1 = k_{\text{initial}}.$$

Determining an appropriate value for k_2 is not as transparent. Most often one would like the above difference equation to converge to a steady state. Hence, one desires that

$$\lim_{t\to\infty} k_t = k_{\text{steady state}},$$

where $k_{\text{steady state}}$ is the long-run steady state. Thus, the goal is to solve the above difference equation subject to two boundary conditions, one at the beginning of time and the other at the end. This falls into a class of problems known as two-point boundary value problems.

Three solution methods for solving this problem are presented. The classic way of solving such problems is multiple shooting. If one knew k_1 and k_2, then the solution for the time path $\{k_t\}_{t=1}^{\infty}$ could be computed by just iterating on (6.1.1). As mentioned, in economics generally only k_1 is known. Multiple shooting selects a value for k_2 such that the economy converges over time to the long-run steady state,

$k_{\text{steady state}}$. Another algorithm is the extended-path method. This algorithm turns the difference equation (6.1.1) on its head. Updating equation (6.1.1) by two periods gives $k_{t+2} = D(k_{t+1}, k_t)$.[1] Rewrite this equation as $k_{t+1} = \widetilde{D}(k_t, k_{t+2})$. For each time period $t = 1, 2, \cdots, \infty$, the extended-path method solves for next period's capital stock, k_{t+1}, given its current value, k_t, and an expectation about its future value two periods down the road, $E[k_{t+2}]$. The algorithm is constructed in such a way so that upon convergence the *expectation* about the *path* for the k_{t+2}'s , or $\{E[k_t]\}_{t=2}$, coincides with the *actual path* for the k_{t+2}'s, or $\{k_t\}_{t=2}$, and also so that the economy converges to a long-run steady state. This is known as a *perfect foresight* equilibrium. The extended-path method and multiple shooting are discussed in Section 6.12. The last solution method treats (6.1.1) either as a second-order linear difference equation that comes out of a linear-quadratic optimization problem or as the linearized Euler equation associated with a nonlinear optimization problem. This method is explained in Section 6.9.

[1] That is, just add 2 to the subscripts in this equation. The updated equation holds for all $t \geq 1$.

The discussion in the chapter will be centered around the neoclassical growth model, which is the workhorse of modern macroeconomics. The model has its roots in work by Frank P. Ramsey (1903-1930). The transitional dynamics for the neoclassical growth model are fully characterized using pencil-and-paper techniques. While doing this, the Bellman (1957) concept of dynamic programming and the value function is presented. Properties of the value function for the neoclassical growth model are derived. The contraction mapping principal underlying much of dynamic programming is discussed. This is done in an intuitive way, at the sacrifice of some rigor. The numerical techniques introduced are illustrated in Section 6.13 using a dynamic version of the monopolist's pricing problem, first introduced in Chapter 2.

6.2 *The World of Robinson Crusoe*

Imagine an economy inhabited by millions of people, all the same. This will be related here in terms of a representative agent, named Robinson Crusoe. Robinson Crusoe's period-t lifetime utility is given by

$$\sum_{j=0}^{\infty} \beta^j U(c_{t+j}), \text{ with } 0 < \beta < 1,$$

where c_{t+j} is the person's consumption in period $t+j$. Utility in period $t+j$, $U(c_{t+j})$, is discounted at the rate β^j. Since $\beta < 1$, β^j is decreasing in j so the further off a utility is in the future the less Robinson cares about it. Note that $\beta^j \to 0$, as $j \to \infty$.

Robinson Crusoe is the name of a famous book written by Daniel Defoe that was first published in 1719. The inscription on the title page read "The Life and Strange Surprizing Adventures of Robinson Crusoe, of York Mariner: Who lived Eight and Twenty Years all alone in an un-inhabited Island of the Coast of America, near the Mouth of the Great River of Oroonoque; Having been cast on shore by Shipwreck, where-in all the Men perished but himself." This is believed to be the first English novel. Interestingly, historians suggest that the book is based on the true story of the buccaneer Alexander Selkirk. After a dispute with his ship's captain, Selkirk was left alone in 1704 on one of the Juan Fernandez Islands for four and a half years. The seaman who went ashore in 1709 to retrieve him said he found "a man clothed in goat'skins, who looked wilder than the first owners of them."

Output in period $t+j$, or o_{t+j}, is produced in line with the following constant-returns-to-scale production function

$$o_{t+j} = \widetilde{F}(k_{t+j}, h_{t+j}),$$

which uses the period-$(t+j)$ inputs, capital, k_{t+j}, and labor, h_{t+j}. To begin with, suppose that Robinson Crusoe supplies just one fixed unit of labor. This restriction will be relaxed in Section 6.10. In light of this restriction, define $F(k_{t+j})$ by $F(k_{t+j}) \equiv \widetilde{F}(k_{t+j}, 1)$. The economy's capital stock is owned by its inhabitants. This capital depreciates at rate δ. Suppose that an individual starts off period $t+j$ owning k_{t+j} units of capital. By investing the amount, i_{t+j}, he can augment the capital stock to k_{t+j+1} in accordance with the law of motion

$$k_{t+j+1} = (1-\delta)k_{t+j} + i_{t+j}.$$

This is a version of Ramsey (1928) growth model.

Frank P. Ramsey (1903-1930) was a British economist and mathematician. He died at the very young age of 26. In economics he is known for his work on the growth model, optimal taxation, and subjective probability. In mathematics he started a branch of combinatorics that is known as Ramsey theory. He was named the Senior Wrangler or the top undergraduate in mathematics at Cambridge.

Robinson Crusoe's goal in life is to maximize his lifetime utility by picking optimally his consumption and investment in each period. His period-t problem can be written as

$$\max_{\{c_{t+j}, i_{t+j}\}_{j=0}^{\infty}} \sum_{j=0}^{\infty} \beta^j U(c_{t+j}),$$

subject to the economy's resource constraint,

$$c_{t+j} + i_{t+j} = \widetilde{F}(k_{t+j}, 1) = F(k_{t+j}),$$

the law of motion for capital,

$$k_{t+j+1} = (1-\delta)k_{t+j} + i_{t+j},$$

and the initial condition, k_t. Robinson Crusoe's problem has been cast as starting in some arbitrary period, t. Often the first period is taken as $t = 1$.[2] Substitute out for c_{t+j} and i_{t+j} in the utility functions using the resource constraint and the law of motion for capital. The problem now appears as

$$\underbrace{V(k_t)}_{\text{VALUE FUNCTION}} \equiv \max_{\{k_{t+j+1}\}_{j=0}^{\infty}} \sum_{j=0}^{\infty} \beta^j U\Big(\underbrace{F(k_{t+j}) + (1-\delta)k_{t+j} - k_{t+j+1}}_{c_{t+j}}\Big). \tag{6.2.1}$$

The function $V(k_t)$ gives the *maximal* level of lifetime utility that Robinson Crusoe will realize if he enters period t with the capital stock, k_t. This is called the *value function*. It plays an important role in modern macroeconomics.

[2] His period-1 problem can be written as

$$\max_{\{c_t, i_t\}_{t=1}^{\infty}} \sum_{t=1}^{\infty} \beta^{t-1} U(c_t),$$

subject to $c_t + i_t = F(k_t)$ and $k_{t+1} = (1-\delta)k_t + i_t$. To see this more formally, set $t = 1$ in the problem in the main text to get

$$\max_{\{c_{1+j}, i_{1+j}\}_{j=0}^{\infty}} \sum_{j=0}^{\infty} \beta U(c_{1+j}),$$

subject to $c_{1+j} + i_{1+j} = F(k_{1+j})$ and $k_{1+j+1} = (1-\delta)k_{1+j} + i_{1+j}$. Now, do a change of variable by setting $t = 1 + j$. Note that if j starts at 0, then t must start at 1. The period-1 problem then obtains.

6.3 *The Euler Equation*

Attention is now directed toward obtaining a solution to Robinson Crusoe's problem (6.2.1). Note that k_{t+j+1} appears exactly twice in the maximand of problem (6.2.1), at time $t+j$ and $t+j+1$. Specifically, it appears in two terms shown below

$$\begin{aligned} ... &+ \beta^j U\Big(F(k_{t+j}) + (1-\delta)k_{t+j} - k_{t+j+1}\Big) \\ &+ \beta^{j+1} U\Big(F(k_{t+j+1}) + (1-\delta)k_{t+j+1} - k_{t+j+2}\Big) + ... \end{aligned}$$

By maximizing with respect to k_{t+j+1}, the following set of first-order conditions can be obtained

$$U_1\Big(F(k_{t+j}) + (1-\delta)k_{t+j} - k_{t+j+1}\Big) = [F_1(k_{t+j+1}) + (1-\delta)] \tag{6.3.1}$$

$$\times \beta U_1\Big(F(k_{t+j+1}) + (1-\delta)k_{t+j+1} - k_{t+j+2}\Big), \text{ for } j = 0, 1, \cdots .$$

This is an infinite dimensional system of equations. The above formula is Robinson Crusoe's *Euler equation*.

Leonhard Euler (1707-1783) is a famous mathematician of Swiss descent. He is known for many things, but his work on fluid dynamics led him to study ordinary and partial differential equations.

The Euler equation plays a central role in modern macroeconomics. It characterizes the consumption/investment decision. The lefthand side represents the cost of investing in an extra unit of capital. Robinson Crusoe must give up one unit of consumption to do this, which has a period $t+j$ utility cost of $U_1(F(k_{t+j}) + (1-\delta)k_{t+j} - k_{t+j+1})$. The righthand side gives the benefit from investing in an extra unit of capital. Output will increase in period $t+j+1$ by the amount $F_1(k_{t+j+1})$. Plus, Crusoe will still have $1-\delta$ units of capital left over after depreciation. An extra unit of period $t+j+1$ consumption is worth $\beta U_1(F(k_{t+j+1}) + (1-\delta)k_{t+j+1} - k_{t+j+2})$ in utility terms. Therefore, investing in an extra unit of capital in period $t+j$ has a utility benefit of $[F_1(k_{t+j+1}) + (1-\delta)] \times \beta U_1(F(k_{t+j+1}) + (1-\delta)k_{t+j+1} - k_{t+j+2})$.

Now,

$$\frac{U_1\Big(\overbrace{F(k_{t+j}) + (1-\delta)k_{t+j} - k_{t+j+1}}^{c_{t+j}}\Big)}{U_1\Big(\underbrace{F(k_{t+j+1}) + (1-\delta)k_{t+j+1} - k_{t+j+2}}_{c_{t+j+1}}\Big)} \gtreqless 1 \text{ as } \beta[F_1(k_{t+j+1}) + (1-\delta)] \gtreqless 1, \text{or as } \beta r_{t+j+1} \gtreqless 1,$$

where $r_{t+j+1} \equiv F_1(k_{t+j+1}) + (1-\delta)$. So, when the gross return on capital accumulation, r_{t+j+1}, exceeds (falls short of) the representative agent's gross rate of subjective time preference, or $1/\beta$, consumption must be growing (falling) over time–the net rate of time preference is defined below. This occurs because $U_1(c_{t+j})/U_1(c_{t+j+1}) > 1$, when $\beta[F_1(k_{t+j+1}) + (1-\delta)] > 1$, which implies that $c_{t+j+1} > c_{t+j}$ due to the assumption of diminishing marginal utility. Think about the gross return on capital accumulation as being equivalent to one plus the market rate of interest, i_{t+j+1}, so that $r_{t+j+1} = 1 + i_{t+j+1}$. Write the gross rate of time preference as $1/\beta = 1 + \rho$, where ρ is the net (or subjective) rate of time preference. Then, consumption grows (declines) when the market interest rate, i_{t+j+1}, exceeds (falls short of) the net rate of time preference, ρ. The market rate of interest represents how much a person will get next period by delaying consumption by one unit. The net rate of time preference speaks to how much an individual requires to postpone a unit of consumption for a period. When $i_{t+j+1} > \rho$ a person will postpone some of their consumption. The above discussion brings up the notion of the elasticity of intertemporal substitution, or how willing a person is to substitute consumption across time in response to changes in the interest rate.

Definition 6.1. (Elasticity of intertemporal substitution) The elasticity of intertemporal substitution is defined as $\iota = -U_1(c_{t+j})/[c_{t+j}U_{11}(c_{t+j})]$. To understand its motivation, consider the consumption Euler equation

$$U_1(c_{t+j}) = \beta r_{t+j} U_1(c_{t+j+1}),$$

where r_{t+j} is the gross interest rate between periods $t+j$ and $t+j+1$. Let r_{t+j} change, while holding fixed c_{t+j+1}, and examine the impact on c_{t+j}.[3] One obtains

$$\frac{dc_{t+j}}{dr_{t+j}} = \frac{\beta U_1(c_{t+j+1})}{U_{11}(c_{t+j})} < 0.$$

This can be expressed in elasticity form as

$$\frac{r_{t+j}}{c_{t+j}} \frac{dc_{t+j}}{dr_{t+j}} = \frac{U_1(c_{t+j})}{c_{t+j}U_{11}(c_{t+j})} = -\iota.$$

Thus, the elasticity of intertemporal substitution controls the response of current consumption to changes in the interest rate.

Example 6.1. (Elasticity of intertemporal substitution) Let $U(c) = c^{1-\sigma}/(1-\sigma)$, with $\sigma \geq 0$, and define the gross real interest, r, by $r \equiv F_1(k) + (1-\delta)$. The Euler equation (6.3.1) reads

$$(\frac{c_{t+j+1}}{c_{t+j}})^{\sigma} = \beta r_{t+j}.$$

Thus, the gross growth rate in consumption, c_{t+j+1}/c_{t+j}, rises with the real interest rate, r_{t+j}. It is easy to calculate that

$$\frac{r_{t+j}}{c_{t+j+1}/c_{t+j}} \frac{d(c_{t+j+1}/c_{t+j})}{dr_{t+j}} = \frac{1}{\sigma},$$

where $1/\sigma$ is the elasticity of intertemporal substitution. The bigger $1/\sigma$ is, the larger the impact that a change in the interest rate, r_{t+j}, has on the growth rate of consumption, c_{t+j+1}/c_{t+j}. As such, the elasticity is important for regulating the responsiveness of consumption and investment to changes in the marginal product of capital. The elasticity of intertemporal substitution will be returned to in Chapter 8.

[3] The term $\beta U_1(c_{t+j+1})$ can be shown to be the value of an extra unit of wealth. Therefore, this experiment is examining the substitution effect of a change in the current interest rate on current consumption, holding the income effect fixed.

The second equates the advantage derived from an increment Δx of consumption at time t, to that derived by postponing it for an infinitesimal period Δt, which will increase its amount to $\Delta x\left(1 + \frac{\partial f}{\partial c}\Delta t\right)$, since $\frac{\partial f}{\partial c}$ gives the rate of interest earned by waiting. This gives

$$u\{x(t)\} = \left\{1 + \frac{\partial f}{\partial c}\Delta t\right\} u\{x(t+\Delta t)\}$$

or in the limit

$$\frac{d}{dt}u(x(t)) = -\frac{\partial f}{\partial c} \cdot u(x(t)) \quad . \quad . \quad . \quad . \quad (3)$$

This equation means that $u(x)$, the marginal utility of consumption, falls at a proportionate rate given by the rate of interest. Consequently x continually increases unless and until either $\frac{\partial f}{\partial c}$ or $u(x)$ vanishes, in which case it is easy to see that bliss must have been attained.

Figure 6.3.1: This is Ramsey (1928) Euler equation. In Ramey's notation $x(t)$ is period-t consumption and $u(x(t))$ is its marginal utility. The marginal product of capital is given by $\partial f/dc$, where f is the production function and c is capital. Ramsey (1928) did not discount the future and capital did not depreciate, which explains the difference between his Euler equation and the modern version (6.3.1). So, in Ramsey's model, the marginal utility of consumption had to decline over time, as stated by his equation (3).

It's interesting to compare equation (6.3.1) with what Ramsey (1928) derived–see Figure 6.3.1.

The Euler equation also represents a 2nd-order nonlinear difference equation, which can be represented implicitly as

$$k_{t+j+2} = D(k_{t+j+1}, k_{t+j}), \text{ for all } j \geq 0.$$

If one knew k_t and k_{t+1} then the above difference equation could be used to solve for k_{t+2} (by setting $j = 0$). One now has k_{t+2} and k_{t+1} in hand that can be used to get k_{t+3}. By iterating on the entire time path for the k_{t+j}'s can be uncovered. That is, one could proceed as follows:

$$\begin{aligned}
k_{t+2} &= D(k_{t+1}, k_t), \\
k_{t+3} &= D(k_{t+2}, k_{t+1}) &&= D\Big(\underbrace{D(k_{t+1}, k_t)}_{k_{t+2}}, k_{t+1}\Big), \\
k_{t+4} &= D(k_{t+3}, k_{t+2}) &&= D\Big(\underbrace{D\big(D(k_{t+1}, k_t), k_{t+1}\big)}_{k_{t+3}}, \underbrace{D(k_{t+1}, k_t)}_{k_{t+2}}\Big), \\
&\vdots &&\vdots
\end{aligned}$$

This procedure generates a sequence for the k_{t+j+2}'s that is a function of the starting values k_t and k_{t+1}–observe that on the righthand side of the above expression things can always be written just in terms of k_t and k_{t+1}, even though the resulting expressions are ugly in form.

Normally, one knows the starting value for capital, k_t. The trouble is that a condition needs to be found to pin down k_{t+1}. Suppose that capital stock converges to the unique steady-state value, k^*. How the steady-state value for the capital stock is determined is discussed next. The steady-state value for the capital stock can be used to tie down k_{t+1}. Specifically, one needs to find the value of k_{t+1} such that $\lim_{j\to\infty} k_{t+j+2} = k^*$. This is called a two-point boundary value problem. The time path is pinned down by an initial condition, k_t, and a terminal condition, k^*. This idea forms the basis of the multiple shooting algorithm discussed in Section 6.12.2.

6.4 *The Steady State*

In a steady state the capital stock will be constant at some level denoted by k^*. Therefore, in a steady state $k_{t+j} = k_{t+j+1} = k_{t+j+2} = \cdots = k^*$. This implies that $c_{t+j} = c_{t+j+1} = c_{t+j+2} = \cdots$ and hence $U_1(c_{t+j}) = U_1(c_{t+j+1}) = U_1(c_{t+j+2}) = \cdots$. So, in this situation the above Euler equation reduces to

$$\beta[F_1(k^*) + (1-\delta)] = 1.$$

Therefore, k^* is determined by

$$F_1(k^*) = 1/\beta - 1 + \delta.$$

Figure 6.4.1 illustrates the situation. Express the discount factor β by $\beta = 1/(1+\iota)$, where $1+\iota$ is the gross rate of time preference and ι is the net rate. Thus, the net rate of time preference is given by $\iota = 1/\beta - 1$. Then, in a steady state

$$F_1(k^*) = \iota + \delta,$$

the marginal product of capital is equal to the (net) rate of time preference, ι, plus the depreciation rate, δ. Since F_1 is a strictly decreasing function of k, this (nontrivial) steady state is *unique*.

Example 6.2. (Steady-state capital stock with a Cobb-Douglas production function) Let production be described by the Cobb-Douglas production function $o = k^\alpha$. Then, $\alpha k^{*\alpha-1} = \iota + \delta$. The steady-state stock of capital, k^*, is then given by $k^* = [\alpha/(\iota+\delta)]^{1/(1-\alpha)}$. The steady-state capital stock declines with the cost of capital, $\iota + \delta$.

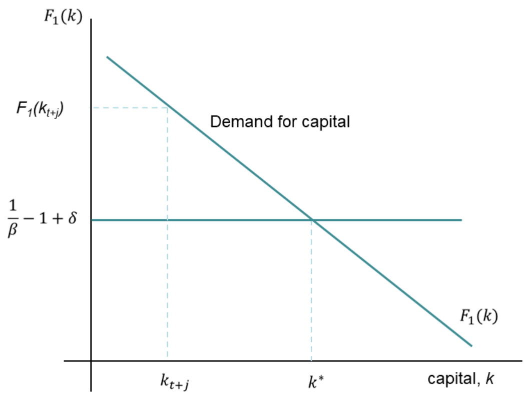


Figure 6.4.1: The diagram illustrates how the steady-state stock of capital, k^*, is determined. The function $F_1(k)$, which specifies the marginal product of capital, is strictly decreasing because the production function, $F(k)$, is strictly concave in the capital stock, k; i.e., there are diminishing marginal returns. This schedule defines the demand for capital. Often the conditions $\lim_{k\to 0} F_1(k) = \infty$ and $\lim_{k\to\infty} F_1(k) = 0$ are imposed. These conditions guarantee that a solution will exist, because $F_1(k)$ must start off above the horizontal $1/\beta - 1 + \delta$ line and end up below. When a solution does exist, it is unique because $F_1(k)$ is downward sloping and can only cross the horizontal $1/\beta - 1 + \delta$ line once. Note that $F_1(k_{t+j}) \gtreqless 1/\beta - 1 + \delta$ as $k_{t+j} \lesseqgtr k^*$.

6.5 Dynamic Programming Formulation

The above optimization problem can be formulated as a *dynamic programming problem*. In problem (6.2.1) there are an infinite number of choice variables, $\{k_{t+j+1}\}_{j=0}^{\infty}$. Bellman (1957) noted that large problems, such as this, suffer from "the curse of dimensionality." His solution was to break down such gigantic problems into a set of smaller simpler problems. In line with this idea, problem (6.2.1) can be recast in terms of a smaller problem for each period $t + j$, which has just one choice variable, k_{t+j+1}. There are effectively an infinite number of these small problems, one for each $t + j$, but they are often easy to compute.

Dynamic programming was introduced in 1953 by the famous applied mathematician Richard E. Bellman (1920-1984) while he was working at the Rand Corporation. It is an important tool in both economics and engineering.

To see this, update problem (6.2.1) by one period (by shifting t to $t + 1$) to get Robinson Crusoe's problem at time $t + 1$:

$$
\begin{aligned}
V(k_{t+1}) &\equiv \max_{\{k_{t+j+2}\}_{j=0}^{\infty}} \sum_{j=0}^{\infty} \beta^j U\Big(F(k_{t+j+1}) + (1-\delta)k_{t+j+1} - k_{t+j+2}\Big) \\
&= \max_{\{k_{t+j+1}\}_{j=1}^{\infty}} \sum_{j=1}^{\infty} \beta^{j-1} U\Big(F(k_{t+j}) + (1-\delta)k_{t+j} - k_{t+j+1}\Big). \quad (6.5.1)
\end{aligned}
$$

Now observe that Robinson Crusoe's time-t problem (6.2.1) can be written as

$$\begin{aligned}V(k_t) &\equiv \max_{\{k_{t+j+1}\}_{j=0}^{\infty}} \sum_{j=0}^{\infty} \beta^j U\Big(F(k_{t+j}) + (1-\delta)k_{t+j} - k_{t+j+1}\Big) \\ &= \max_{k_{t+1}} \Big\{ U\Big(F(k_t) + (1-\delta)k_t - k_{t+1}\Big) \\ &+ \max_{\{k_{t+j+1}\}_{j=1}^{\infty}} \Big\{ \beta \sum_{j=1}^{\infty} \beta^{j-1} U\Big(F(k_{t+j}) + (1-\delta)k_{t+j} - k_{t+j+1}\Big)\Big\}\Big\} \\ &= \max_{k_{t+1}} \Big\{ U\Big(F(k_t) + (1-\delta)k_t - k_{t+1}\Big) + \beta V(k_{t+1})\Big\} \\ &\text{[using (6.5.1)]}. \end{aligned} \tag{6.5.2}$$

Note how k_{t+1} can be separated from the inner maximization problem. This can only be done since the return function $U(F(k_t) + (1 - \delta)k_t - k_{t+1})$ doesn't involve future values of the control variable, here $\{k_{t+j+1}\}_{j=1}^{\infty}$. This allows the maximization to be separated into two maximization operations, with the max operator in the outer problem cascading over the one in the inner problem.

The agent's period-t dynamic programming problem is

$$V(k_t) = \max_{k_{t+1}} \Big\{ U\Big(F(k_t) + (1-\delta)k_t - k_{t+1}\Big) + \beta V(k_{t+1})\Big\}.$$

This is called the *Bellman equation*. (To get the period $t + j$ problem just rewrite t as $t + j$). Effectively, this is just a two-period problem; viz, today and the future. The future is encapsulated in the function $V(k_{t+1})$. This function gives the maximal level of lifetime utility that can be obtained in period $t + 1$ contingent on Robinson having the capital stock k_{t+1}.

The first-order condition for optimality in the period-t is

$$U_1\Big(F(k_t) + (1-\delta)k_t - k_{t+1}\Big) = \beta V_1(k_{t+1}). \tag{6.5.3}$$

This will determine his consumption-savings decision. The lefthand side of this equation is the marginal cost associated with doing an extra unit of investment in period t. An extra unit of investment reduces consumption by one which in turn reduces period-t utility by $U_1(F(k_t) + (1 - \delta)k_t - k_{t+1})$. The righthand side is the marginal benefit. In particular, $V_1(k_{t+1})$ in the increase in lifetime utility from having an extra unit of capital in period $t + 1$. Since this gain is one period down the road it should discounted by β. Equation (6.5.3) represents one equation in one unknown, k_{t+1}. The solution to the above first-order condition has the form

$$k_{t+1} = K(k_t).$$

The function K is known as a decision rule. This is a first-order difference equation. Some properties of this decision rule are discussed further below. The steady-state capital stock, k^*, must satisfy the condition

$$k^* = K(k^*).$$

The discussion now turns to establishing two properties of the value function, $V(k_t)$. First, it will be shown that $V(k_t)$ is strictly increasing in k_t. Second, it will be demonstrated that $V(k_t)$ is strictly concave in k_t.

6.5.1 *Differentiation of the Value Function using the Envelope Theorem*

Next, differentiate both sides of (6.5.2) to get

$$\begin{aligned} V_1(k_t) &= U_1\Big(F(k_t) + (1-\delta)k_t - k_{t+1}\Big)[F_1(k_t) + (1-\delta)] \\ &\quad + \underbrace{[-U_1\Big(F(k_t) + (1-\delta)k_t - k_{t+1}\Big) + \beta V_1(k_{t+1})]}_{=0} \times dk_{t+1}/dk_t \\ &= U_1\Big(F(k_t) + (1-\delta)k_t - k_{t+1}\Big)[F_1(k_t) + (1-\delta)] > 0. \end{aligned} \tag{6.5.4}$$

The term in the middle disappears because of the first-order condition for the maximization. This is called the *envelope theorem*. (For a simple explanation of the envelope theorem, see Chapter A.) The result implies that $V(k_t)$ is increasing in the capital stock. So, not surprisingly, Robinson Crusoe is better off in period t the more capital, k_t, he has. This result is stated now as a lemma.

Lemma 6.1. *(Value function is strictly increasing) $V(k_t)$ is strictly increasing.*

The above expression can be updated to period $t+1$ by changing the time subscripts on the variables. Doing this gives

$$V_1(k_{t+1}) = U_1\Big(F(k_{t+1}) + (1-\delta)k_{t+1} - k_{t+2}\Big)[F_1(k_{t+1}) + (1-\delta)].$$

Using this on the righthand side of (6.5.3) yields

$$U_1\Big(F(k_t) + (1-\delta)k_t - k_{t+1}\Big) = \beta U_1\Big(F(k_{t+1}) + (1-\delta)k_{t+1} - k_{t+2}\Big)[F_1(k_{t+1}) + (1-\delta)],$$

or equivalently, by rewriting t as $t+j$ as,

$$\begin{aligned} &U_1\Big(F(k_{t+j}) + (1-\delta)k_{t+j} - k_{t+j+1}\Big) \\ &= \beta U_1\Big(F(k_{t+j+1}) + (1-\delta)k_{t+j+1} - k_{t+j+2}\Big)[F_1(k_{t+j+1}) + (1-\delta)]. \end{aligned}$$

This is the Euler equation (6.3.1) again. So, the two approaches yield the same solution to optimization problem.

6.5.2 Concavity of the Value Function

Last, the function $V(k_t)$ is strictly concave.

Definition 6.2. (Strict concavity) A function $V : \mathcal{K}^n \rightarrow \mathcal{R}$ is strictly concave if

$$V(\theta k^1 + (1-\theta)k^2) > \theta V(k^1) + (1-\theta)V(k^2),$$

for all $k^1, k^2 \in \mathcal{K}^n$ such that $k^1 \neq k^2$ and $\theta \in (0,1)$. A function $V : \mathcal{K}^n \rightarrow \mathcal{R}$ is concave if $V(\theta k^1 + (1-\theta)k^2) \geq \theta V(k^1) + (1-\theta)V(k^2)$, for all $k^1, k^2 \in \mathcal{K}^n$ such that $k^1 \neq k^2$ and $\theta \in (0,1)$. Note in general that k can be a n-dimensional vector, even though this case is not analyzed here. Figure 6.5.1 illustrates the definition for the situation where $n = 1$.

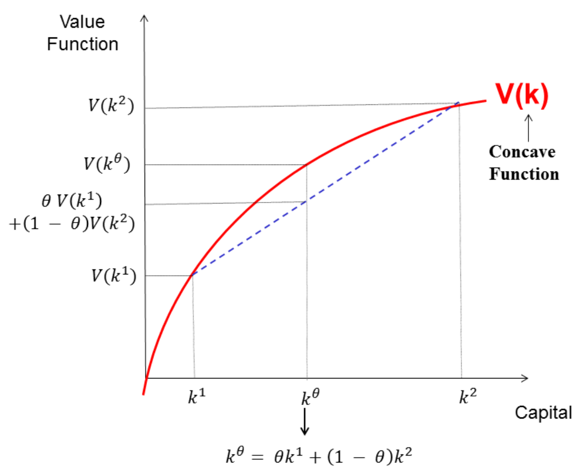


Figure 6.5.1: The figure illustrates how $V(k^\theta) > \theta V(k^1) + (1-\theta)V(k^2)$ when $V(k)$ is strictly concave. Let the equation for the dashed straight line be $V(k) = a + bk$. It is easy to see that $\theta V(k^1) = \theta a + \theta b k^1$ and $(1-\theta)V(k^2) = (1-\theta)a + (1-\theta)bk^2$. Hence, $\theta V(k^1) + (1-\theta)V(k^2) = a + bk^\theta$, as shown. Although irrelevant, $a = [V(k^1)k^2 - V(k^2)k^1]/(k^2 - k^1)$ and $b = [V(k^1) - V(k^2)]/(k^2 - k^1)$.

Lemma 6.2. *(Value function is strictly concave) $V(k_t)$ is strictly concave.*

Proof. Consider two points, viz k_t^1 and k_t^2. Take a convex combination of these two points; let $k_t^\theta = \theta k_t^1 + (1-\theta)k_t^2$, for $\theta \in (0,1)$. Need to show that $\theta V(k_t^1) + (1-\theta)V(k_t^2) < V(k_t^\theta)$. Now, let $k_{t+j}^\theta = \theta k_{t+j}^{*1} + (1-\theta)k_{t+j}^{*2}$, where k_{t+j}^{*1} and k_{t+j}^{*2} are the optimal solutions for k_{t+j}^1 in (6.2.1) starting off from the initial condition $k_t = k_t^1$ and $k_t = k_t^2$, respectively. (So, to be clear, in this proof an asterisk does not refer to the steady-state value for capital.) It will be shown that the k_{t+j}^θ's are feasible solutions when starting off from k_t^θ. Then,

$$\begin{aligned}\theta V(k_t^1) + (1-\theta)V(k_t^2) &= \theta \max_{\{k_{t+j+1}^1\}_{j=0}^\infty} \sum_{j=0}^\infty \beta^j U\Big(F(k_{t+j}^1) + (1-\delta)k_{t+j}^1 - k_{t+j+1}^1\Big) \\ &+ (1-\theta) \max_{\{k_{t+j+1}^2\}_{j=0}^\infty} \sum_{j=0}^\infty \beta^j U\Big(F(k_{t+j}^2) + (1-\delta)k_{t+j}^2 - k_{t+j+1}^2\Big) \\ &= \theta \sum_{j=0}^\infty \beta^j U\Big(F(k_{t+j}^{*1}) + (1-\delta)k_{t+j}^{*1} - k_{t+j+1}^{*1}\Big) \\ &+ (1-\theta) \sum_{j=0}^\infty \beta^j U\Big(F(k_{t+j}^{*2}) + (1-\delta)k_{t+j}^{*2} - k_{t+j+1}^{*2}\Big).\end{aligned}$$

Now concavity of the utility and production functions on the right-hand side, in addition to assuming feasibility (shown below), allows this to be rewritten as

$$\theta V(k_t^1) + (1-\theta)V(k_t^2) < \sum_{j=0}^\infty \beta^j U\Big(F(k_{t+j}^\theta) + (1-\delta)k_{t+j}^\theta - k_{t+j+1}^\theta\Big).$$

Last note that on the righthand side of the above expression k_{t+j}^θ and k_{t+j+1}^θ are not optimal so that

$$\theta V(k_t^1) + (1-\theta)V(k_t^2) < \sum_{j=0}^\infty \beta^j U\Big(F(k_{t+j}^{*\theta}) + (1-\delta)k_{t+j}^{*\theta} - k_{t+j+1}^{*\theta}\Big) = V(k_t^\theta).$$

It needs to be shown that k_{t+j}^θ is a feasible solution. Note that if $0 < k_{t+j+1}^1 < F(k_{t+j}^1) + (1-\delta)k_{t+j}^1$ and $0 < k_{t+j+1}^2 < F(k_{t+j}^2) + (1-\delta)k_{t+j}^2$, then $0 < k_{t+j+1}^\theta < F(k_{t+j}^\theta) + (1-\delta)k_{t+j}^\theta$, by the concavity of F. Therefore, $\{k_{t+j+1}^\theta\}_{j=0}^\infty$ is a feasible solution for the problem associated with $V(k_t^\theta)$, even though it may not be optimal. □

So, before proceeding on to analyzing the transitional dynamics for the neoclassical growth model, it has been shown that the value function, $V(k_t)$, is strictly increasing, and strictly concave in k_t.

6.6 *Consumption Smoothing*

Imagine that Robinson Crusoe gets a tiny bit more capital in period t. This will increase his period-t resources by $F_1(k_t) + (1-\delta)$. One would expect that he will consume some of this windfall increase in resources and that he will save the rest. That is, one might expect that

$$0 < K_1(k_t) < F_1(k_t) + (1-\delta). \tag{6.6.1}$$

I.e., that fact that $K_1(k_t) > 0$ implies that Robinson must be saving some of the windfall increase in capital, while the condition $K_1(k_t) < F_1(k_t) + (1-\delta)$ means that he must be consuming part of it.

Property (6.6.1) can be derived in two ways. The first approach uses the first-order condition connected with the dynamic programming problem (6.5.2). Update (6.5.3) to get

$$U_1\Big(F(k_{t+j}) + (1-\delta)k_{t+j} - k_{t+j+1}\Big) = \beta V_1(k_{t+j+1}). \tag{6.6.2}$$

By totally differentiating this equation it can be seen that

$$\begin{aligned} 0 < \frac{dk_{t+j+1}}{dk_{t+j}} &= \underbrace{\frac{U_{11}(\cdot_{t+j})}{U_{11}(\cdot_{t+j}) + \beta V_{11}(\cdot_{t+j+1})}}_{<1}[F_1(\cdot_{t+j}) + 1 - \delta] \\ &< F_1(\cdot_{t+j}) + 1 - \delta. \end{aligned} \tag{6.6.3}$$

The notation $\cdot_{t+j}$ signifies that the arguments of the function are being evaluated at time $t+j$. Note that the above derivation assumes that the value function is continuously twice differentiable; there may be some points where this is not the case. Equation (6.6.3) implies that period-$(t+j)$ consumption, c_{t+j}, will rise because $dk_{t+j+1}/dk_{t+j} < F_1(\cdot_{t+j}) + 1 - \delta$. So, the representative agent is both consuming and saving (since $0 < dk_{t+j+1}/dk_{t+j}$) some of income resulting from an increase in the period-$(t+j)$ capital stock. Now, by updating the above formula it follows that $0 < dk_{t+j+2}/dk_{t+j+1} < F_1(\cdot_{t+j+1}) + 1 - \delta$. Therefore, the increase in k_{t+j+1} will also cause both c_{t+j+1} and k_{t+j+2} to rise. And, the increase in k_{t+j+2} will induce c_{t+j+2} and k_{t+j+3} to move up and so on. So, the effect of increase in the period-$(t+j)$ capital stock will propagate throughout the entire future increasing consumption and the capital stock in every period.

The second approach employs the Euler equation (6.3.1).[4] To show that property (6.6.1) is consistent with the Euler equation (6.3.1), substitute out for k_{t+j+2} using the updated relationship $k_{t+j+2} = K^i(k_{t+j+1})$ to get

[4] The second approach can be skipped for those not interested in formalities.

$$\begin{aligned} U_1\Big(F(k_{t+j}) + (1-\delta)k_{t+j} - k_{t+j+1}\Big) &= \beta[F_1(k_{t+j+1}) + (1-\delta)] \\ &\quad \times U_1\Big(F(k_{t+j+1}) + (1-\delta)k_{t+j+1} - K^i(k_{t+j+1})\Big). \end{aligned}$$

Here $K^i(k_{t+j+1})$ is some guess for $K(k_{t+j+1})$. This is one equation in one unknown, k_{t+j+1}. As an induction hypothesis at stage i, suppose that

$$0 \le K_1^i \le F_1 + 1 - \delta. \tag{6.6.4}$$

It will now be shown that this implies

$$0 < \frac{dk_{t+j+1}}{dk_{t+j}} < F_1 + 1 - \delta.$$

Totally differentiate the above Euler equation to get

$$\frac{dk_{t+j+1}}{dk_{t+j}} = \frac{U_{11}(\cdot_{t+j})[F_1(\cdot_{t+j}) + 1 - \delta]}{\Delta} = \frac{U_{11}(\cdot_{t+j})}{\Delta}[F_1(\cdot_{t+j}) + 1 - \delta],$$

where

$$\Delta \equiv U_{11}(\cdot_{t+j}) + \beta F_{11}(\cdot_{t+j+1})U_1(\cdot_{t+j+1})$$
$$+ \beta[F_1(\cdot_{t+j+1}) + (1-\delta)]U_{11}(\cdot_{t+j+1})[F_1(\cdot_{t+j+1}) + 1 - \delta - K_1^i(\cdot_{t+j+1})].$$

Then, it is easy to see that

$$0 < \frac{dk_{t+j+1}}{dk_{t+j}} = K_1^{i+1} < F_1 + 1 - \delta,$$

because $U_{11}(\cdot_{t+j})/\Delta < 1$. Hence, the property is self-fulfilling. Now, $\lim_{i\to\infty} K_1^i = K$. Thus,

$$0 < \frac{dk_{t+j+1}}{dk_{t+j}} = K_1 < F_1 + 1 - \delta.$$

6.7 *Dynamics*

The dynamics of the neoclassical growth model are developed now. The analysis starts off with Lemma 6.3 that states if one starts off in the current period $t+j$ with a capital stock, k_{t+j}, that lies below the steady-state capital stock, k^*, then next period's capital stock, k_{t+j+1}, will be bigger than the current one. This will be useful for showing monotonic convergence toward the steady-state capital stock.

Lemma 6.3. *(If below the steady state, then rise, while if above, then fall.) If $k_{t+j} < k^*$, then $k_{t+j} < k_{t+j+1}$, and if $k_{t+j} > k^*$, then $k_{t+j} > k_{t+j+1}$.*

Proof. Recall that the value function is strictly concave so that $V_1(k_{t+j})$ is strictly decreasing in k_{t+j}. Therefore, $V_1(k_{t+j}) \gtreqless V_1(k_{t+j+1})$ as $k_{t+j} \lesseqgtr k_{t+j+1}$–see Figure 6.7.1. Hence,

$$[V_1(k_{t+j}) - V_1(k_{t+j+1})](k_{t+j} - k_{t+j+1}) < 0.$$

By using the envelope theorem, it was shown that $V_1(k_{t+j}) = U_1\Big(F(k_{t+j}) + (1-\delta)k_{t+j} - k_{t+j+1}\Big)[F_1(k_{t+j}) + (1-\delta)]$–this is an updated version of (6.5.4). The first-order condition for k_{t+j+1}, or equation (6.6.2), also implied that $V_1(k_{t+j+1}) = U_1\Big(F(k_{t+j}) + (1-\delta)k_{t+j} - k_{t+j+1}\Big)/\beta$. Plugging these two expressions into the above condition gives

$$[F_1(k_{t+j}) + (1-\delta) - 1/\beta](k_{t+j} - k_{t+j+1}) < 0.$$

Suppose $k_{t+j} < k^*$. Then, $F_1(k_{t+j}) + (1-\delta) - 1/\beta > 0$–see Figure 6.4.1. But, this implies $k_{t+j} - k_{t+j+1} < 0$ or $k_{t+j+1} > k_{t+j}$. Conversely, if $k_{t+j} > k^*$ then $k_{t+j} < k_{t+j+1}$. □

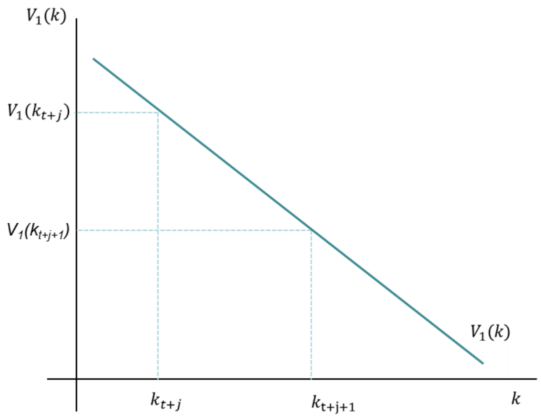


Figure 6.7.1: The graph illustrates that if $k_{t+j} < k_{t+j+1}$, then $V_1(k_{t+j}) > V_1(k_{t+j+1})$. This transpires because the value function, $V(k)$, is strictly concave in the capital stock, k, implying that $V_1(k) < 0$.

The dynamics for the Ramsey growth model can now be characterized, with Lemma 6.3 in hand. They are portrayed in Figure 6.7.2. From (6.6.3) the decision rule for capital, $k_{t+j+1} = K(k_{t+j})$, is strictly increasing. Along the 45 degree line $k_{t+j+1} = k_{t+j}$. Hence, a steady state is located at points where the K function crosses this line. There can only be one nontrivial steady state, as was discussed above. When k_{t+j} is below k^* the function $K(k_{t+j})$ lies above the 45 degree line, by the Lemma 6.3. In this situation, the function $K(k_{t+j})$ cannot return a value for k_{t+j+1} greater than k^*. If it did, then the function K would have to turn down to attain the steady state, which can't happen because K is strictly increasing. When k_{t+j} is above k^* the function $K(k_{t+j})$ lies below the 45 degree line. It must cut the 45 degree line from above due to the Lemma 6.3. This implies that at the steady state

$$0 < \frac{dk_{t+j+1}}{dk_{t+j}} < 1, \tag{6.7.1}$$

because the slope of the 45 degree line is one. A local solution for dk_{t+j+1}/dk_{t+j} around the nontrivial steady state is given in Section 9.5. This is obtained by linearizing the Euler equation (6.3.1) around the (unique nontrivial) steady state. It will be reaffirmed then that (6.7.1) holds by examining the linear difference equation that arises from the linearized Euler equation.

To conclude, the model's transitional dynamics are as displayed by Figure 6.7.2. When starting off below the steady state the capital stock monotonically increases until it converges to its steady-state value. Along the transition path toward the steady state, the interest rate steadily falls. To see this, note that the period-$(t+j)$ gross interest rate is given by $U_1(F(k_{t+j}) + (1-\delta)k_{t+j} - k_{t+j+1})/[\beta U_1(F(k_{t+j+1}) + (1-\delta)k_{t+j+1} - k_{t+j+2})] = F_1(k_{t+j+1}) + 1 - \delta$. The term on the left is the amount of period-$(t+j+1)$ consumption that the person must

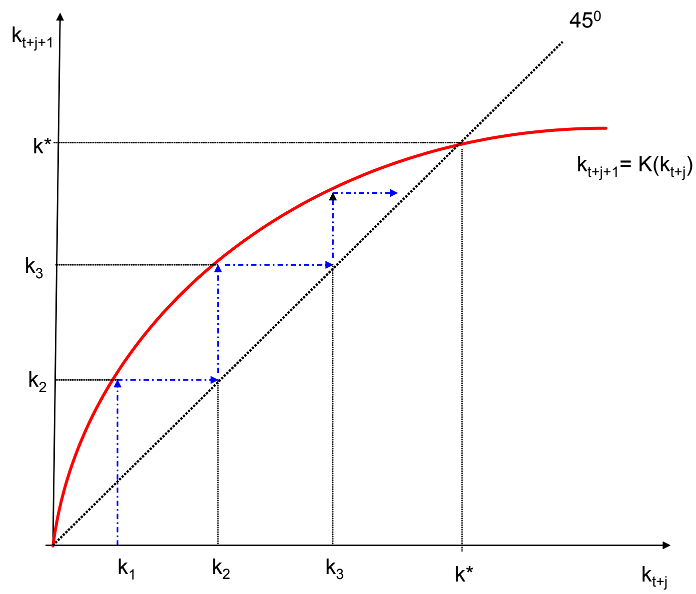


Figure 6.7.2: Transitional Dynamics. The economy starts off in period 1 with the capital stock k_1, displayed on the horizontal axis. Using the decision rule, $k_{t+j+1} = K(k_{t+j})$, this implies that the capital stock in period 2 will be k_2, as shown on the vertical axis. To move forward in time to period 2, reflect the period-2 capital stock, k_2, onto the horizontal axis using the 45^o line. It can then be seen, by using the decision rule again, that the period-3 capital stock will be k_3, as given on the vertical axis. The capital stock will keep raising in a monotone fashion until the steady-state value of capital, k^*, is reached.

receive in order to sacrifice a unit of period-$(t+j)$ consumption. This is set equal to the gross return from investing in capital in period $t+j$, or the term on the right. Clearly, the term on the right decreases over time as the capital stock increases. The story is reversed when starting off from above the steady state.

Situations such as those shown in Figure 6.7.3 are ruled out by the uniqueness property. If a second (non-trivial) steady state did exist (which it does not), then it would have to be unstable. At the second steady state the policy function cuts the 45 degree line from below implying $K_1 > 1$. The above lemma established that this can't happen. To see why, observe from Figure 6.7.3 that in a neighborhood around the unstable k^* if $k_{t+j} > k^*$ then $k_{t+j+1} > k_{t+j}$, which would contradict the lemma. Note that the trivial steady state in Figure 6.7.2 is unstable. If the system is started from a value for k_{t+j} that is close to zero it will always converge to k^* and not zero.

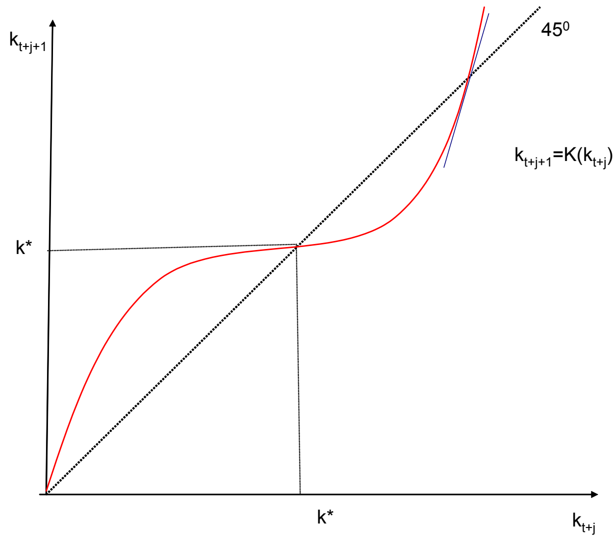


Figure 6.7.3: Unstable equilibria. If at any steady state $dk_{t+j+1}/dk_{t+j} < 1$, then an unstable nontrivial steady state cannot exist.

6.8 The Value Function: A More Formal Analysis

The above analysis suggests that the neoclassical growth model can be written as

$$V(k) \equiv \max_{k'}\{U(F(k) + (1-\delta)k - k') + \beta V(k')\}. \qquad \text{P(1)}$$

The messy time subscripts have been eliminated in the above dynamic programming problem where next period's capital stock has a prime symbol attached to it. This can be done because there is no notion of time in Robinson Crusoe's problem. All that matters is the amount of capital that he enters a period with. The goal is to answer the following questions concerning the value function, V:

1. Will V exist?
2. Is V unique?
3. Is V continuous?
4. Is V continuously differentiable?
5. Is V increasing in k?
6. Is V concave in k?

6.8.1 Method of Successive Approximation

The idea here is to approximate the value function V by a sequence of successively better guesses, denoted by V^j at stage j. Consider the following algorithm to do this:

1. Make an initial guess for V. Call it V^0.

2. Construct a revised guess for V, denoted by V^1:

$$V^1(k) \equiv \max_{k'}\{U(F(k) + (1-\delta)k - k') + \beta V^0(k')\}.$$

3. Enter iteration $n+1$ with a solution for V from the previous iteration, V^n. Compute V^{n+1}, given V^n, as follows

$$V^{n+1}(k) \equiv \max_{k'}\{U(F(k) + (1-\delta)k - k') + \beta V^n(k')\}. \qquad \text{P(2)}$$

This procedure can be represented much more compactly using *operator* notation.

$$V^{n+1} = TV^n.$$

The operator T is shorthand notation for the list of operations, described by P(2), which are performed on the function V^n to transform it into the new one V^{n+1}. Often the operator T maps some set of functions, say $\mathcal{F}$, into itself. That is, $T : \mathcal{F} \to \mathcal{F}$. The hope is that as n gets large it will transpire that $V^n \to V$, where $V = TV$. To know if V^n is close to V requires some sort of metric or a standard for measuring distance. This brings up the notion of a metric space.

6.8.2 *Metric Spaces: A Detour through Real Analysis*

Definition 6.3. (Metric Space) A metric space is a set $\mathcal{S}$, together with a metric $\rho : \mathcal{S} \times \mathcal{S} \to \mathcal{R}_+$, such that for all $x, y, z \in \mathcal{S}$ (see Figure 6.8.1):

1. $\rho(x,y) \geq 0$, with $\rho(x,y) = 0$ if and only if $x = y$,

2. $\rho(x,y) = \rho(y,x)$,

3. $\rho(x,z) \leq \rho(x,y) + \rho(y,z)$.

The points in a metric space can actually be functions. On this, think about a continuous function $x(n)$ as just being an infinite dimensional vector; i.e., the infinite dimensional analogue of the point $x = \{x_1, \cdots, x_j, \cdots, x_n\}$ in $\mathcal{R}^n$ where now j can vary continuously. How can the distance between two continuous function be measured?

Definition 6.4. (Uniform Metric) Consider the space of continuous functions $\mathcal{C} : [a,b] \to \mathcal{R}$. A useful metric for this space is

$$\rho(x,y) = \max_{t\in[a,b]} |x(t) - y(t)|,$$

where $x(t)$ and $y(t)$ are two functions in $\mathcal{C}$. This is called the uniform metric.

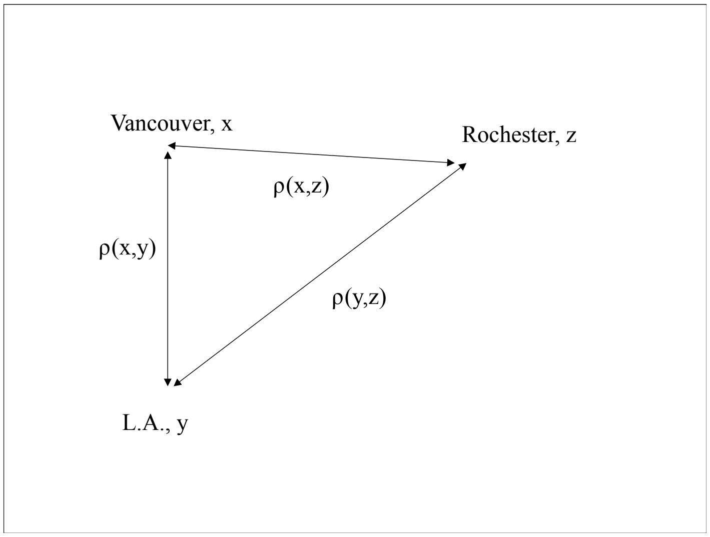


Figure 6.8.1: Distances Between Cities. The distance between Rochester and Vancouver is non-negative. The miles from Rochester to Vancouver are the same as the from Vancouver to Rochester. Taking a detour through L.A. increases the miles covered.

Example 6.3. (Distance between two functions–uniform metric) Figure 6.8.2 plots the two continuous functions $x(t) = 1$ and $y(t) = 1 + t - t^2$ on the space $[0, 1]$. When using the uniform metric the functions are farthest apart at the point $t = 0.5$.

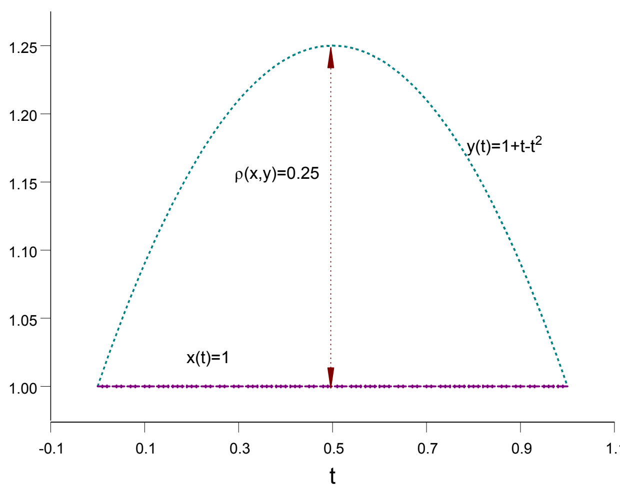


Figure 6.8.2: The Uniform Metric: The maximal distance between the two function $x(t)$ and $y(t)$ occurs at the point $t = 0.5$.

The iterative scheme P(2) generates a sequence of functions $\{V^n\}_{n=0}^{\infty}$. Will this sequence converge to something? What does convergence mean?

Definition 6.5. (Convergence of a Sequence) A sequence $\{x_n\}_{n=0}^{\infty}$ in $\mathcal{S}$ converges to $x \in \mathcal{S}$, if for each $\varepsilon > 0$ there exists a N_ε such that

$$\rho(x_n, x) < \varepsilon, \text{ for all } n \geq N_\varepsilon.$$

In general N_ε will depend on ε. From a computational viewpoint this notation of convergence isn't very appealing. Suppose one is trying to find a numerical solution, $x \in \mathcal{S}$, to some problem. To get the solution a computer algorithm is employed. The algorithm generates a sequence $x_0, x_1, x_2, \ldots$. Each point in the sequence is hopefully getting closer and closer to the answer x. When should one stop? The above criteria is not be very useful to use because it requires knowing the answer x, and this is what is being sought.

Definition 6.6. (Cauchy Sequence) A sequence $\{x_n\}_{n=0}^{\infty}$ in $\mathcal{S}$ is a Cauchy sequence if for each $\varepsilon > 0$ there exists a N_ε such that

$$\rho(x_m, x_n) < \varepsilon, \text{ for all } m, n \geq N_\varepsilon.$$

From the computational viewpoint the Cauchy criteria for convergence looks more appealing. Basically, it would say keep iterating until the answers being generated aren't changing much. Of course, it would be impossible to check whether or not $\rho(x_m, x_n) < \varepsilon$ for *all* $m, n \geq N_\varepsilon$. Also, will a Cauchy sequence generated by some algorithm converge to an answer, $x \in \mathcal{S}$? The answer in general is no.

Remark 6.1. A Cauchy sequence in $\mathcal{S}$ may not converge to a point in $\mathcal{S}$.

Example 6.4. (A Cauchy sequence in $\mathcal{S}$ that converges to point outside of $\mathcal{S}$) Let $\mathcal{S} = (0, 1]$, $\rho(x, y) = |x - y|$, and $\{x_n\}_{n=1}^{\infty} = \{1/n\}_{n=1}^{\infty}$. Clearly, $x_n \to 0 \notin (0, 1]$. This sequence satisfies the Cauchy criteria, though, because

$$\rho(x_n, x_m) = |\frac{1}{m} - \frac{1}{n}| \leq \frac{1}{m} + \frac{1}{n} < \varepsilon, \text{ if } m, n > \frac{2}{\varepsilon}.$$

Given this it is often useful to focus attention on those metric spaces $(\mathcal{S}, \rho)$ where all Cauchy sequences are guaranteed to converge to a point in the space.

Definition 6.7. (Complete Metric Space) A metric space $(\mathcal{S}, \rho)$ is complete if every Cauchy sequence in $\mathcal{S}$ converges to a point in $\mathcal{S}$.

Theorem 6.1. *Let $X \subseteq \mathcal{R}^l$ and $\mathcal{C}(X)$ be the set of bounded continuous functions $V : X \to \mathcal{R}$ with the uniform metric $\rho(V, W) = \sup_{x \in X} |V - W|$. Then $\mathcal{C}(X)$ is a complete metric space.*

Proof. See Bryant (1985, Theorem 3.9). □

Remark 6.2. Pointwise convergence of a sequence of continuous functions does not imply that the limiting function is continuous.

Example 6.5. (Pointwise converge of a sequence of continuous functions to a discontinous function) Let $\{V^n\}_{n=1}^{\infty}$ in $\mathcal{C}[0, 1]$ be defined by

$V^n(t) = t^n$. As $n \to \infty$ it transpires that: (i) $V^n(t) \to 0$ for $t \in [0,1)$ and (ii), $V^n(t) \to 1$ for $t = 1$. Thus,

$$V(t) = \begin{cases} 0, & \text{for } t \in [0,1), \\ 1, & \text{for } t = 1. \end{cases}$$

Hence $V(t)$ is a discontinuous function. See Figure 6.8.3. Clearly, by the above theorem $\{V^n\}_{n=1}^{\infty}$ cannot describe a Cauchy sequence under the uniform metric. This can be shown directly too, however. In particular, for given *any* N_ε it is always possible to pick a $m, n \geq N_\varepsilon$ and $t \in [0,1)$ so $|t^n - t^m| \geq 1/2$. To see this pick $n = N_\varepsilon$ and a $t \in (0,1)$ so that $t^{N_\varepsilon} \geq 3/4$; i.e., choose $t \geq (3/4)^{1/N_\varepsilon}$. Next, pick m large enough such that, for the t chosen earlier, $t^m < 1/4$ or $m \geq (\ln 1/4)/(\ln t)$. The desired results obtains.

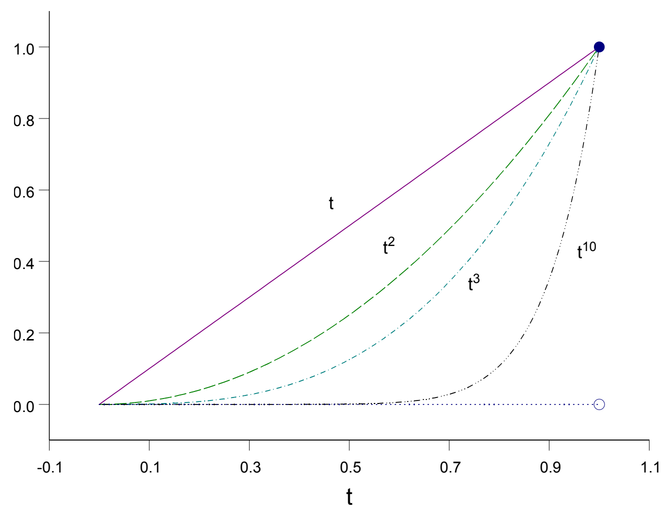


Figure 6.8.3: Pointwise Convergence of Continuous Functions to a Discontinuous Function. As n increases the continuous functions $V^n(t) = t^n$ bend more. Eventually the stress is too much and the limiting function, $V(t)$, breaks at the point $t = 1$.

Remark 6.3. The space of strictly increasing functions is not complete since the limiting function may just be nondecreasing. Likewise, the space of strictly concave functions is not complete since the limiting function may just be concave.

Example 6.6. Consider the Cauchy sequence of strictly increasing, strictly concave function $\{y = x^{1-1/(n+1)}\}_{n=1}^{\infty}$ on the domain $[0,1]$. This sequence converges to the increasing, concave function $y = x$. The situation is shown in Figure 6.8.4.

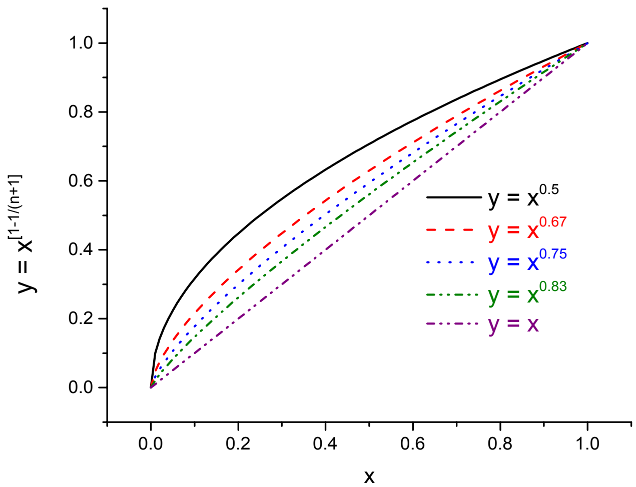


Figure 6.8.4: The Spaces of Strictly Concave and Strictly Increasing Functions are Not Complete. For example, the sequence of strictly concave, strictly increasing functions shown converges to a straight line, which is just a concave, strictly increasing function.

6.8.3 *The Contraction Mapping Theorem*

Establishing, both computationally and theoretically, properties of mappings such as P(2) involves the idea of a contraction mapping.

Definition 6.8. (Contraction Mapping) Let $(\mathcal{S}, \rho)$ be a metric space and $T : \mathcal{S} \to \mathcal{S}$ be function mapping $\mathcal{S}$ into itself. T is a contraction mapping (with modulus β) if for $\beta \in (0, 1)$,

$$\rho(Tx, Ty) \leq \beta\rho(x, y), \text{ for all } x, y \in \mathcal{S}. \tag{6.8.1}$$

As the name implies, after applying the operator T the distance between functions contracts; i.e., the distance between Tx and Ty is smaller than between x and y.

Theorem 6.2. *(Contraction Mapping Theorem or Banach Fixed Point Theorem) If $(\mathcal{S}, \rho)$ is a complete metric space and $T : \mathcal{S} \to \mathcal{S}$ is a contraction mapping with modulus β, then*

1. T has exactly one fixed point $V \in \mathcal{S}$ such that $V = TV$,
2. for any $V^0 \in \mathcal{S}$, $\rho(T^n V^0, V) \leq \beta^n \rho(V^0, V)$, $n = 0, 1, 2, \ldots$.

Proof. See Bryant (1985, Theorem 4.1). □

The theorem implies that from a computational standpoint contraction mappings are great. Consider the mapping $V^{n+1} = TV^n$, where the operator T is a contraction. Part 1 of the theorem states that there is only one fixed point to the operator. Part 2 says that you can get to this unique fixed point by employing the iterative scheme $V^{n+1} = TV^n$ starting from *any* initial guess V^0 (in the space $\mathcal{S}$).

Corollary 6.1. *Let $(\mathcal{S}, \rho)$ be a complete metric space and let $T : \mathcal{S} \to \mathcal{S}$ be a contraction mapping with fixed point $V \in \mathcal{S}$. If $\mathcal{S}'$ is a closed subset of $\mathcal{S}$ and $T(\mathcal{S}') \subseteq \mathcal{S}'$ then $V \in \mathcal{S}'$. If in addition $T(\mathcal{S}') \subseteq \mathcal{S}'' \subseteq \mathcal{S}'$, then $V \in \mathcal{S}''$.*

Proof. Choose $V^0 \in \mathcal{S}'$ and note that $\{T^n V^0\}$ is a sequence in $\mathcal{S}'$ converging to V. Since $\mathcal{S}'$ is closed, it follows that $V \in \mathcal{S}'$. If $T(\mathcal{S}') \subseteq \mathcal{S}''$, it then follows that $V = TV \in \mathcal{S}''$. □

To check whether a particular mapping is a contraction using (6.8.1) can be cumbersome. So, for dynamic programming problems in economics it is often much easier to use the sufficient conditions presented below.

David Blackwell (1919-2010) was an American mathematician and statistician. He made important contributions to game theory, information theory, probability theory, and statistics. He was the first African American to become a tenured professor at Berkeley and the first to be inducted into the National Academy of Sciences.

Theorem 6.3. *(Blackwell's Sufficiency Condition) Let $\mathcal{X} \subseteq \mathcal{R}^l$ and $\mathcal{B}(X)$ be the space of bounded functions $V : \mathcal{X} \to \mathcal{R}$ with the uniform metric. Let $T : \mathcal{B}(\mathcal{X}) \to \mathcal{B}(\mathcal{X})$ be an operator satisfying*

1. (Monotonicity) $V, W \in \mathcal{B}(\mathcal{X})$. If $V \leq W$ [i.e., $V(x) \leq W(x)$ for all x] then $TV \leq TW$.
2. (Discounting) There exists some constant $\beta \in (0, 1)$ such that $T(V + a) \leq TV + \beta a$, for all $V \in \mathcal{B}(\mathcal{X})$ and $a \geq 0$.

Then T is a contraction with modulus β.

Proof. For every $V, W \in \mathcal{B}(\mathcal{X})$,

$$V \leq W + \rho(V, W).$$

Thus, (1) and (2) imply

$$TV \leq \underbrace{T(W + \rho(V, W))}_{\text{Monotonicity}} \leq \underbrace{TW + \beta\rho(V, W)}_{\text{Discounting}}.$$

Thus,

$$TV - TW \leq \beta\rho(V, W).$$

By permuting the functions it is easy to show that

$$TW - TV \leq \beta\rho(V, W).$$

Consequently,

$$|TV - TW| \leq \beta\rho(V, W),$$

so that

$$\rho(TV, TW) \leq \beta\rho(V, W).$$

Therefore T is a contraction. □

6.8.4 *Back to the Neoclassical Growth Model*

The above machinery will now be applied to the neoclassical growth model. To this end, consider the mapping

$$(TV)(k) = \max_{0 \le k' \le F(k)+(1-\delta)} \{U(F(k) + (1-\delta)k - k') + \beta V(k')\}. \quad \text{P(3)}$$

Is T a contraction?

1. *Monotonicity*. Suppose $V(k) \le W(k)$ for all k. It will be shown that $(TV)(k) \le (TW)(k)$.

$$(TV)(k) = \{U(F(k) + (1-\delta)k - k'^{*}) + \beta V(k'^{*})\},$$

   where k'^{*} maximizes P(3). Clearly,

$$\begin{aligned}(TV)(k) &\le \{U(F(k) + (1-\delta)k - k'^{*}) + \beta W(k'^{*})\} \\ &\le \max_{0 \le k' \le F(k)+(1-\delta)} \{U(F(k) + (1-\delta) - k') + \beta W(k')\} \\ &= (TW)(k).\end{aligned}$$

2. *Discounting*.

$$\begin{aligned}T(V+a)(k) &= \max_{0 \le k' \le F(k)+(1-\delta)} \{U(F(k) + (1-\delta)k - k') + \beta[V(k') + a]\} \\ &= \max_{0 \le k' \le F(k)+(1-\delta)} \{U(F(k) + (1-\delta)k - k') + \beta V(k')\} + \beta a \\ &= (TV)(k) + \beta a.\end{aligned}$$

Theorem 6.4. *V is a continuous, strictly increasing, strictly concave function in k.*[5]

[5] A more formal proof is in Stokey and Lucas (1986).

Proof. (Heuristic) It will be shown that the operator described by P(3) maps increasing, concave $\mathcal{C}^2$ functions into strictly increasing, strictly concave $\mathcal{C}^2$ functions. Suppose that V^n is a continuous, strictly increasing, strictly concave $\mathcal{C}^2$ function. The decision rule for k' is determined from the first-order condition

$$U_1(F(k) + (1-\delta)k - k') = \beta V_1^n(k').$$

This determines k' as a continuously differentiable function of k by the implicit function theorem. Therefore, $V^{n+1}(k)$ is a strictly increasing $\mathcal{C}^2$ function, since $V_1^{n+1}(k) = U_1(F(k) + (1-\delta)k - k')F_1(k) > 0$. The limit of such a sequence must be a continuous function, because each V^n is a continuous function and the sequence converges uniformly. (It is does *not* have to be a $\mathcal{C}^2$ function) The limiting function is also strictly increasing because P(3) maps increasing functions into strictly increasing ones.[6] To see this, let $k_1 < k_2$. Then,

[6] In terms of Collorary 6.1 think about $\mathcal{S}'$ as being space of increasing functions and $\mathcal{S}''$ the space of strictly increasing functions.

$$\begin{aligned} V^{n+1}(k_1) &= U(F(k_1) + (1-\delta)k_1 - k_1'^*) + \beta V^n(k_1'^*) \\ &< U(F(k_2) + (1-\delta)k_2 - k_1'^*) + \beta V^n(k_1'^*) \\ &\leq U(F(k_2) + (1-\delta)k_2 - k_2'^*) + \beta V^n(k_2'^*) \\ &= V^{n+1}(k_2). \end{aligned}$$

By employing a proof similar to that used in Lemma 6.2 it can be shown that the operator described by P(3) maps concave function into strictly concave ones.[7] Thus, the limiting function must be strictly concave. □

[7] Again, in terms of Collorary 6.1 think about $\mathcal{S}'$ as being space of concave functions and $\mathcal{S}''$ the space of strictly concave functions.

Differentiability

The last question concerns whether or not the value function for the neoclassical growth model is differentiable.

Lemma 6.4. *Let $X \subseteq \mathcal{R}^l$ be a convex set, $V : X \to \mathcal{R}$ be a concave function. Pick an $x_0 \in \text{INT}(X)$ and let D be a neighborhood of x_0. If there is a concave, differentiable function $W : D \to \mathcal{R}$ with $W(x_0) = V(x_0)$ and $W(x) \leq V(x)$ for all $x \in D$ then V is differentiable at x_0 and*

$$V_i(x_0) = W_i(x_0), \text{ for } i = 1, 2, ..., l.$$

Proof. (Heuristic) Figure 6.8.5 tells it all. If V is not differentiable at x_0, then it would have to have a kink in it at this point. But, if this was the case it would be impossible to have a smooth function W lying always below V that just touches V at x_0. Try to derive a contradiction by drawing different scenarios. □

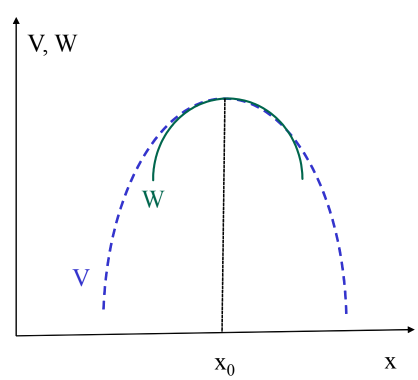


Figure 6.8.5: Differentiability of V. The function V cannot have a kink (or a break in its derivative) at the point x_0. If it did, then it would be impossible to insert the concave, differentiable function W under the function V while touching at x_0.

Theorem 6.5. *(Benveniste and Scheinkman) Suppose that K is a convex set and that U and F are strictly concave C^1 functions. Let $V : K \to \mathcal{R}$ in line with P(3) and denote the decision rule associated with this problem by $k' = G(k)$. Pick $k_0 \in \text{INT}(K)$ and assume that $0 < G(k_0) < F(k_0) + (1-\delta)k_0$. Then $V(k)$ is continuously differentiable at k_0 with its derivative given by*

$$V_1 = U_1\big(F(k_0) + (1-\delta)k_0 - G(k_0)\big)[F_1(k_0) + (1-\delta)k_0].$$

Proof. Clearly, there exists some neighborhood D of k_0 such that $0 < G(k_0) < F(k) + (1-\delta)k$ for all $k \in D$. Define W on D by

$$W(k) = U\big(F(k) + (1-\delta)k - G(k_0)\big) + \beta V\big(G(k_0)\big).$$

Now, W is concave and differentiable since U and F are. Furthermore, it follows that

$$W(k) \leq \max_{k'}\{U\big(F(k) + (1-\delta)k - k'\big) + \beta V(k')\} = V(k),$$

with this expression holding with strict equality at $k = k_0$. The results then follow immediately from the above lemma. □

6.9 A Linear-Quadratic Optimization Problem

Robinson Crusoe's problem is now recast as a linear-quadratic optimization problem. This class of optimization problems is characterized by a quadratic objective function and linear constraints. They yield linear first-order conditions and are a close cousin of the linearization technique discussed in Chapter 9. The solution to the linear-quadratic optimization problem will shed further light on the local dynamics of the neoclassical growth model.

6.9.1 Taking a Quadratic Approximation to the Utility Function

Substitute the resource constraint into the momentary utility function to obtain

$$U(F(k) + (1-\delta)k - k').$$

Take a second-order Taylor expansion of this to get

$$\begin{aligned} U(F(k) + (1-\delta)k - k') = U(*) &+ \underbrace{U_1(*)[F_1(*) + (1-\delta)]}_{\alpha}(k-k^*) - \underbrace{U_1(*)}_{\lambda}(k'-k^*) \\ &+ \frac{1}{2}\underbrace{[U_{11}(*)[F_1(*)+(1-\delta)]^2 + U_1(*)F_{11}(*)]}_{-\psi}(k-k^*)^2 \\ &- \underbrace{U_{11}(*)[F_1(*)+(1-\delta)]}_{-\rho}(k-k^*)(k'-k^*) + \frac{1}{2}\underbrace{U_{11}(*)}_{-\phi}(k'-k^*)^2. \end{aligned}$$

(See Chapter A for the concept of a second-order Taylor expansion.) In the above equation the $*$ notation signifies that the argument of the function is being evaluated at its steady-state value. Define deviations in capital stock from the steady state by $\widehat{k} \equiv k - k^*$ and $\widehat{k}' \equiv k' - k^*$. The momentary utility function can then be expressed as

$$U(\widehat{k},\widehat{k}') = \tau + \alpha\widehat{k} - \lambda\widehat{k}' + \rho\widehat{k}\widehat{k}' - \frac{\psi}{2}\widehat{k}^2 - \frac{\phi}{2}\widehat{k}'^2.$$

Note the slight abuse of notation in redefining U to now be a function of $\widehat{k}$ and $\widehat{k}'$. The derivatives in the above Taylor expansion can be computed numerically using the formulae presented for numerical first- and second-derivatives presented in Chapter 8.

6.9.2 *Robinson's Linear-Quadratic Optimization Problem*

Robinson Crusoe's optimization problem is now given by

$$\max_{\{k_{t+j+1}\}_{j=0}^{\infty}} \sum_{j=0}^{\infty} \beta^j U(k_{t+j}, k_{t+j+1}), \text{ with } 0 < \beta < 1.$$

As before, k_{t+j+1} will show up twice in the objective function:

$$\cdots + \beta^j U(k_{t+j}, k_{t+j+1}) + \beta^{j+1} U(k_{t+j+1}, k_{t+j+2}) + \cdots$$

Maximizing then gives

$$U_2(k_{t+j}, k_{t+j+1}) = -\beta U_1(k_{t+j+1}, k_{t+j+2}), \text{ for } j = 0, 1, \cdots.$$

The Euler equation for capital accumulation is then given by

$$\underbrace{\overbrace{\lambda - \rho\widehat{k}_{t+j} + \phi\widehat{k}_{t+j+1}}^{-U_2(k_{t+j},k_{t+j+1})}}_{\text{MC of investment}} = \underbrace{\overbrace{\beta(\alpha + \rho\widehat{k}_{t+j+2} - \psi\widehat{k}_{t+j})}^{=U_1(k_{t+j+1},k_{t+j+2})}}_{\text{MB of investment}},$$

or

$$\lambda - \rho\widehat{k} + \phi\widehat{k}' = \beta(\alpha + \rho\widehat{k}'' - \psi\widehat{k}').$$

This is a linear second-order difference equation.

Now, in a steady state $\widehat{k}^* = \widehat{k}^{*\prime} = \widehat{k}^{*\prime\prime} = 0$, so that the following parameter restriction must apply:

$$\lambda = \beta\alpha.$$

Therefore,

$$-\rho\widehat{k} + \phi\widehat{k}' = \beta(\rho\widehat{k}'' - \psi\widehat{k}').$$

Conjecture a solution of the form

$$\widehat{k}' = \eta\widehat{k},$$

so that

$$\widehat{k}'' = \eta \widehat{k}'.$$

There is no constant term since the decision rule has been defined in terms of deviations from the steady state. Note that if η was negative, then the capital stock would oscillate around the steady state, which Figure 6.7.2 rules out. So, one should expect that $\eta > 0$. Now, if $0 < \eta < 1$, then the difference equation is stable and convergence to the steady state will be monotone.

Using the conjectured decision rule in the above Euler equation gives

$$-\rho \widehat{k} + \phi \widehat{k}' = \beta(\rho \eta \widehat{k}' - \psi \widehat{k}'),$$

which can be rewritten as

$$\widehat{k}' = \frac{\rho}{(\phi + \beta\psi - \beta\rho\eta)} \widehat{k}.$$

Hence, η must solve the quadratic

$$\eta = \frac{\rho}{\phi + \beta\psi - \beta\rho\eta}.$$

Cross multiplying gives

$$-\beta\rho\eta^2 + (\phi + \beta\psi)\eta - \rho = 0.$$

This equation will have two roots. One will lie between 0 and 1 and the other will be greater than 1. The stable root corresponds to the situation where the decision rule crosses the 45^0 degree line in Figure 6.7.2.

To prove this formally, observe that when $\eta = 0$ the lefthand side of this equation is negative. When $\eta = 1$ then the lefthand side is positive because

$$\begin{aligned}
\beta\psi &= -U_{11}(*)[F_1(*) + (1-\delta)] - \beta U_1(*) F_{11}(*), \\
\beta\rho &= -U_{11}(*), \\
\phi &= -U_{11}(*), \\
\rho &= -U_{11}(*)[F_1(*) + (1-\delta)],
\end{aligned}$$

so that

$$\begin{aligned}
-\beta\rho + (\phi + \beta\psi) - \rho = \\
& U_{11}(*) - U_{11}(*) - U_{11}(*)[F_1(*) + (1-\delta)] \\
& - \beta U_1(*) F_{11}(*) + U_{11}(*)[F_1(*) + (1-\delta)] \\
& = -\beta U_1(*) F_{11}(*) > 0.
\end{aligned}$$

Therefore, a root must lie between 0 and 1. As η becomes large the first term in the quadratic equation will dominate and the expression turns negative again. Figure 6.9.1 portrays the situation. The transitional dynamics for the neoclassical growth model are revisited in Chapter 8.

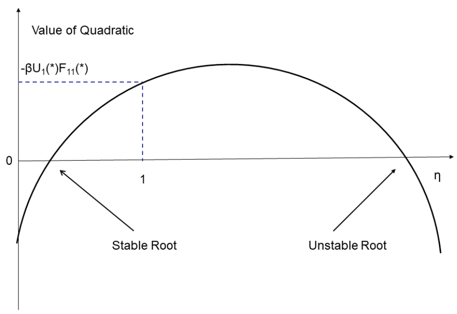


Figure 6.9.1: Roots to the quadratic equation for η. The unstable root can be thrown away, implying $0 < \eta < 1$. Thus, the difference equation $\widehat{k}' = \eta\widehat{k}$ will be stable and exhibit monotone dynamics.

6.10 *Adding a Labor-Leisure Choice*

It is easy to add a labor-leisure choice to the above analysis. To do this, rewrite the period-$(t+j)$ momentary utility function as

$$U\Big(c_{t+j} - G(h_{t+j})\Big),$$

where h_{t+j} is the amount of work effort expended. This utility function was introduced in Chapter 2. The function $G : \mathcal{R}_+ \to \mathcal{R}_+$ gives the disutility of work effort, measured in consumption units. Assume that it is strictly increasing and strictly convex. Thus, the utility cost of each extra unit of work effort rises with level of work. Express the production function as

$$o_{t+j} = F(k_{t+j}, h_{t+j}).$$

Robinson Crusoe's problem will now appear as

$$\max_{\{h_{t+j}, k_{t+j+1}\}_{j=0}^{\infty}} \sum_{j=0}^{\infty} \beta^j U\Big(F(k_{t+j}, h_{t+j}) + (1-\delta)k_{t+j} - k_{t+j+1} - G(h_{t+j})\Big).$$

There will be two first-order conditions; viz, one for h_{t+j} and the other for k_{t+j+1}. The first-order condition governing period-$(t+j)$ labor effort, or h_{t+j}, is

$$\begin{aligned} &U_1\Big(F(k_{t+j}, h_{t+j}) + (1-\delta)k_{t+j} - k_{t+j+1} - G(h_{t+j})\Big)F_2(k_{t+j}, h_{t+j}) \\ &\quad = U_1\Big(F(k_{t+j}, h_{t+j}) + (1-\delta)k_{t+j} - k_{t+j+1} - G(h_{t+j})\Big)G_1(h_{t+j}), \end{aligned}$$

which simplifies to

$$F_2(k_{t+j}, h_{t+j}) = G_1(h_{t+j}).$$

This equation specifies h_{t+j} as a function of k_{t+j}. Write this solution as

$$h_{t+j} = H(k_{t+j}).$$

The first-order condition for the period-$(t+j+1)$ capital stock, k_{t+j+1}, is

$$U_1\Big(F(k_{t+j},h_{t+j})+(1-\delta)k_{t+j}-k_{t+j+1}-G(h_{t+j})\Big)$$
$$=[F_1(k_{t+j+1},h_{t+j+1})+(1-\delta)]$$
$$\times\,\beta U_1\Big(F(k_{t+j+1},h_{t+j+1})+(1-\delta)k_{t+j+1}-k_{t+j+2}-G(h_{t+j+1})\Big),$$

for $j=0,1,\cdots$. Plugging in the decision-rule for labor then yields

$$U_1\left(F\big((k_{t+j},H(k_{t+j})\big)+(1-\delta)k_{t+j}-k_{t+j+1}-G\big(H(k_{t+j})\big)\right)$$
$$=[F_1\left(k_{t+j+1},H(k_{t+j+1})\right)+(1-\delta)]$$
$$\times\,\beta U_1\Big(F\big(k_{t+j+1},H(k_{t+j+1})\big)+(1-\delta)k_{t+j+1}-k_{t+j+2}-G\big(H(k_{t+j+1})\big)\Big),$$

for $j=0,1,\cdots$. The problem has been reduced to the earlier one; i.e., the solution for the model is once again represented by a second-order nonlinear difference equation for the capital stock.

Example 6.7. (Hours worked with a zero-income effect utility function and a Cobb-Douglas production function) Let $G(h)=h^{1+\theta}/(1+\theta)$ and $F(k,h)=k^{\alpha}h^{1-\alpha}$. Then, the first-order condition for labor can be written as

$$\underbrace{(1-\alpha)k^{\alpha}h^{-\alpha}}_{F_2}=\underbrace{h^{\theta}}_{G_1}.$$

It's trivial to calculate that

$$h=[(1-\alpha)k^{\alpha}]^{1/(\theta+\alpha)}.$$

6.11 *The Taxation of Capital and Labor Income*

Adding income taxation into the above framework with a labor choice is straightforward. This can be done along the lines discussed in Chapter 2. To see how, let labor income be taxed at the rate τ_h. Likewise, let the tax rate on capital income, net of the cost of depreciation, be τ_k. For simplicity assume that all tax revenue is rebated back to the consumer/worker in the form of lump-sum transfer payments, λ.[8]- Suppose that consumer/workers own all the capital in the economy, which they rent out to firms at the rate r. Likewise, they supply labor at the wage rate w. Assume that tastes and technology have the forms given in the previous section.

[8] Footnote 9 outlines how to add government spending.

The representative consumer/worker's period-$(t+j)$ budget constraint reads

$$c_{t+j}+k_{t+j+1}=(1-\tau_h)w_{t+j}h_{t+j}+r_{t+j}k_{t+j}-\tau_k(r_{t+j}-\delta)k_{t+j}+(1-\delta)k_{t+j}+\lambda_{t+j}$$
$$=(1-\tau_h)w_{t+j}h_{t+j}+[1+(1-\tau_k)(r_{t+j}-\delta)]k_{t+j}+\lambda_{t+j}.$$

Observe that the person only pays taxes on their per unit rental income, r_{t+j}, net of depreciation, δ; i.e., their tax on capital income is $\tau_k(r_{t+j} - \delta)k_{t+j}$. As before, this equation can be used to substitute out for c_{t+j} in the consumer/worker's utility function. The government's budget constraint will appear as

$$\tau_h w_{t+j} h_{t+j} + \tau_k(r_{t+j} - \delta)k_{t+j} = \lambda_{t+j}.$$

When formulating the general equilibrium for this economy, do the following steps in order:

1. Solve the consumer/worker's and firm's problems. This will lead to an Euler equation for capital accumulation of the form

$$U_1\Big(c_{t+j} - G(h_{t+j})\Big) = \beta[1 + (1-\tau_k)(r_{t+j+1} - \delta)] \times U_1\Big(c_{t+j+1} - G(h_{t+j+1})\Big),$$

   where c_{t+j} and c_{t+j+1} are given by the consumer/worker's constraint. The first-order condition for labor is

$$(1 - \tau_h)w_{t+j} = G_1(h_{t+j}).$$

2. Eliminate λ_{t+j} in the consumer/worker's budget constraint using the government's budget constraint. This will result in

$$c_{t+j} + k_{t+j+1} = w_{t+j}h_{t+j} + r_{t+j}k_{t+j} + (1-\delta)k_{t+j}.$$

3. Use the firm's first-order conditions to solve out for the wage and rental rates in consumer/worker's formula for consumption. After employing Euler's theorem, this will result in[9]

$$c_{t+j} = F(k_{t+j}, h_{t+j}) + (1-\delta)k_{t+j} - k_{t+j+1}.$$

A set of equations will arise that only involve the k's and h's. To make the computer code readable, this line could be inserted before the Euler equation presented in Step 1. Similarly, when h_{t+j} has an analytical solution, it could be placed before the line for c_{t+j}. The expressions for c_{t+j+1} and h_{t+j+1} can be obtained by just updating the formulas for c_{t+j} and h_{t+j}.

Return to concept of the equivalent variation that was introduced in Chapter 2 and consider a switch from some tax regime A to tax regime B. How much would a person be willing to pay, as a fraction of each period's consumption under regime A, to move from A to B? The fraction ϵ solves the equation

$$\sum_{j=0}^{\infty} \beta^j U\Big(c^A_{t+j}(1+\epsilon) - G(h^A_{t+j})\Big) = W^B \equiv \sum_{j=0}^{\infty} \beta^j U\Big(c^B_{t+j} - G(h^B_{t+j})\Big).$$

Things are a little more complicated now, but this is just one equation in the unknown variable ϵ.

[9] If government spending is added into the mix, then the government's budget constraint will read

$$\tau_h w_{t+j} h_{t+j} + \tau_k(r_{t+j} - \delta)k_{t+j} = \lambda_{t+j} + g_{t+j}.$$

This will result in the equation for consumption appearing as

$$c_{t+j} = F(k_{t+j}, h_{t+j}) + (1-\delta)k_{t+j} - k_{t+j+1} - g_{t+j}.$$

6.12 The Extended Path and Multiple Shooting Algorithms

Unlike the quadratic case, in general the solution to the neoclassical growth model must be solved numerically on a computer. The above framework will be modified to allow for some predetermined known sequence of technology shocks, $\{z_t\}_{t=1}^{\infty}$. In particular, assume that Robinson Crusoe's time-1 choice problem is now given by

$$\max_{c_t, k_{t+1}} \sum_{t=1}^{\infty} \beta^{t-1} U(c_t),$$

subject to

$$c_t + k_{t+1} = F(k_t, z_t) + (1-\delta)k_t,$$

and the initial condition, k_1. Note the presence of the technology shock in the production function. The Euler equation for this model is

$$\begin{aligned} &U_1\Big(F(k_t, z_t) + (1-\delta)k_t - k_{t+1}\Big) \\ &= \beta U_1\Big(F(k_{t+1}, z_{t+1}) + (1-\delta)k_{t+1} - k_{t+2}\Big)[F_1(k_{t+1}, z_{t+1}) + (1-\delta)]. \end{aligned} \tag{6.12.1}$$

Assume that the economy converges to the steady-state level of the capital stock by period T. Let z^* denote the steady-state level for the technological shock. The steady-state level of capital, k^*, will then be given by

$$1/\beta = F_1(k^*, z^*) + 1 - \delta. \tag{6.12.2}$$

So, a time path for capital is being sought that goes from k_1 to $k_T = k^*$. Thus, essentially a solution is being sought for $T-2$ capital stocks, or for $k_2, k_3, \cdots, k_{T-1}$. When simulating the neoclassical growth model, it is a good idea not to set the terminal period, T, too large. This will be elaborated on when discussing the multiple shooting algorithm.

6.12.1 The Extended Path Algorithm

The extended path algorithm was proposed by Fair and Taylor (1983). Observe that if k_{t+2} is known, then (6.12.1) can be used to solve for k_{t+1}, given k_t. This observation suggests the following algorithm:

1. Enter iteration j with a guess for the sequence $\{k_t\}_{t=1}^{T}$, denoted by $\{k_t^j\}_{t=1}^{T}$. For each period t (for $t = 1, 2, \cdots, T-2$) solve for k_{t+1} using the equation

$$U_1\Big(\underbrace{F(k_t, z_t) + (1-\delta)k_t - k_{t+1}}_{c_t}\Big) = \beta U_1\Big(\underbrace{F(k_{t+1}, z_{t+1}) + (1-\delta)k_{t+1} - k_{t+2}^j}_{c_{t+1}}\Big) \times [F_1(k_{t+1}, z_{t+1}) + (1-\delta)]. \tag{6.12.3}$$

Note that k_t was determined in the previous period and that k^j_{t+2} is given by the guess. This will generate a sequence $\{k_t\}_{t=1}^T$. Specifically, using a "for" or "do" loop:

(a) Start off in period 1 with the predetermined starting condition, k_1. Use (6.12.3) to determine k_2, given k_1 and the guess k^j_3.

(b) Move to period 2. Here the goal is to calculate k_3, given the solution just obtained for k_2 together with the guess k^j_4.

(c) Then compute k_4, given k_3 and k^j_5.

(d) Proceed down the time path in the above fashion to period $T-2$. Here the starting capital stock is k_{T-2}. Finally, compute k_{T-1}, given the guess $k^j_T = k_T = k^*$.

2. Check whether $\sum_{t=1}^T |k_t - k^j_t| < \varepsilon$.

(a) If so, exit the algorithm since a solution has been found.

(b) If not, set $\{k^{j+1}_t\}_{t=1}^T = \{k_t\}_{t=1}^T$. Repeat step one using this new guess.

Extended Path Algorithm

Time, t	Variables			
	Predetermined		Computed	Future Guess
1	k_1		k_2	k^j_3
		$\swarrow$		
2	k_2		k_3	k^j_4
		$\swarrow$		
3	k_3		k_4	k^j_5
$\vdots$	$\vdots$		$\vdots$	$\vdots$
$T-2$	k_{T-2}		k_{T-1}	$k^j_T = k^*$

Remark 6.4. (Restricting consumption to be positive) It pays sometimes to impose a lower bound on consumption to avoid the overshooting that often occurs with Newton's method. In particular, one could add the lines $c_t = F(k_t, z_t) + (1-\delta)k_t - k_{t+1}$, $c_t = \max\{1.0E-5, c_t\}$, $c_{t+1} = F(k_{t+1}, z_{t+1}) + (1-\delta)k_{t+1} - k_{t+2}$, and $c_{t+1} = \max\{1.0E-5, c_{t+1}\}$ just before the Euler equation is presented and write the marginal utilities as $U_1(c_t)$ and $U_1(c_{t+1})$.

(Speeding up Newton's Method) Newton's method can be sped up by using the solution for k_{t+1} on iteration $j-1$ as the starting guess for the nonlinear equation solver on iteration j. That is, use k^j_{t+1} as the starting guess in the solver for k_{t+1}. The same is true if the Bisection method is used where a guess for the solution is provided, instead of specifying upper and lower bounds.

6.12.2 *Multiple Shooting*

Again suppose that the economy converges to the steady-state level of the capital stock, k^*, by period T. Let z^* denote the steady-state level for the technological shock. Once again k^* will be defined by (6.12.2). Recall that the Euler equation (6.12.1) is a second-order difference equation in the capital stock. Thus, two starting conditions are needed, k_1 and k_2. Now, k_1 is predetermined. So, the idea of the algorithm is to pick k_2 so that the capital stock will be k^* at time T. This can be expressed in terms of finding the solution to one nonlinear equation in one unknown variable, k_2.

1. At the heart of the algorithm is constructing a function that returns a value for the terminal capital stock, k_T, given the two starting values, k_1 and k_2, and the sequence of technology shocks, $\{z_t\}_{t=1}^{\infty}$. Denote this function by $k_T = K_T(k_1, k_2; z_1, \cdots, z_{T-1})$. Suppose one has a guess for the capital stock in period 2, denoted by k_2^j. The period-1 stock of capital, k_1, is a predetermined variable. Given a guess for k_2, denoted by k_2^j, a value for k_3 can be computed using the Euler equation for capital accumulation. Call this a guess for k_3, or k_3^j. The idea is to solve for the sequence of capital stocks $\{k_t^j\}_{t=3}^{T}$ using the Euler equation for capital accumulation. That is, given guesses for k_t and k_{t+1}, denoted by k_t^j and k_{t+1}^j, one can solve recursively for $k_{t+2}^j, k_{t+3}^j, \cdots, k_T^j$, using the second-order nonlinear difference equation

$$U_1\left(F(k_t^j, z_t) + (1-\delta)k_t^j - k_{t+1}^j\right) = \beta U_1\left(F(k_{t+1}^j, z_{t+1}) + (1-\delta)k_{t+1}^j - k_{t+2}^j\right) \times [F_1(k_{t+1}^j, z_{t+1}) + (1-\delta)]. \quad (6.12.4)$$

The above difference equation implicitly generates a sequence for the capital stocks described by the following:

$$k_3^j = D(k_2^j, k_1, z_2, z_1) \equiv K_3(k_1, k_2^j; z_1, z_2),$$
$$k_4^j = D(k_3^j, k_2^j, z_3, z_2) = D\left(D(k_2^j, k_1, z_2, z_1), k_2^j, z_3, z_2\right) \equiv K_4(k_1, k_2^j; z_1, z_2, z_3)$$
$$\vdots$$
$$k_T^j = K_T(k_1, k_2^j; z_1, \cdots, z_{T-1}).$$

This ultimately gives a value for k_T^j that is effectively based on k_1^j and k_2^j. Since the first capital stock is predetermined so that $k_1^j = k_1$. this solution can be represented by

$$k_T^j = K_T(k_1, k_2^j; z_1, \cdots, z_{T-1}).$$

On the computer, the function K_T will involve writing a "for" or "do" loop as follows:

(a) Start off in period 3. Given the starting value k_1 and the guess k_2^j one gets k_3^j.

(b) Move on to period 4. Given k_2^j and k_3^j one can obtain k_4^j. and so on.

(c) Proceed down the time path in the above manner until one gets to period T. Here one will enter the period with k_{T-2}^j and k_{T-1}^j. The goal is to solve for the final capital stock, k_T.

2. At time T it is desired that $k_T^j \simeq k^*$, the steady-state capital stock. So, the algorithm amounts to solving the following nonlinear equation for k_2,

$$K_T(k_1, k_2; z_1, \cdots, z_{T-1}) - k^* = 0, \tag{6.12.5}$$

where the function $K_T(k_1, k_2; z_1, \cdots, z_{T-1})$ is characterized in Step 1. This nonlinear equation can be solved by either bisection or Newton's method. Either method essentially involves iterating on the starting condition, k_2^j, until $|k_T^j - k^*| < \varepsilon$, where the nonlinear equation is given by (6.12.5).

(a) When a value for k_2^j is found that sets $|k_T^j - k^*| < \varepsilon$, the nonlinear equation solver will terminate since a solution has been found.

(b) If not, the nonlinear equation solver will try a new guess for k_2, denoted by k_2^{j+1}. It will repeat step one using this new guess.

MULTIPLE SHOOTING ALGORITHM

Find the Zero of $K_T(k_1, k_2; z_1, \cdots, z_{T-1}) - k^* = 0$ via choice of k_2

	Variables	
Predetermined	Guess	Target
k_1	k_2	$K_T(k_1, k_2; z_1, \cdots, z_{T-1}) = k^*$

Loop inside Function, $K_T(k_1, k_2; z_1, \cdots, z_{T-1})$

Time, t	Variables				
		Predetermined			Computed
3	k_1		k_2		k_3
		$\swarrow$		$\swarrow$	
4	k_2		k_3		k_4
		$\swarrow$		$\swarrow$	
5	k_3		k_4		k_5
$\vdots$	$\vdots$		$\vdots$		$\vdots$
T	k_{T-2}		k_{T-1}		k_T

The idea of the algorithm is to pick the starting value for the period-2 capital stock so that the second-order difference equation $k_{t+2} =$

$D(k_{t+1}, k_t, z_{t+1}, z_t)$ converges to the steady state at the end of T periods. Now, if $k_2^j \neq k_2$ the computed sequence of k's will tend to diverge away in an explosive manner from the true solution.

To see this, suppose that economy starts off from a value for k_1 that lies in some small neighborhood of the final steady state. Now, make a guess for k_2 that lies above the true value. By using (6.12.4) when $t = 1$, it is easy to calculate that

$$\begin{aligned}\frac{dk_3}{dk_2} &= \frac{U_{11}(\cdot_1) + \beta[F_1(\cdot_2) + (1-\delta)]^2 U_{11}(\cdot_2) + \beta F_{11}(\cdot_2) U_1(\cdot_2)}{\beta[F_1(\cdot_2) + (1-\delta)]U_{11}(\cdot_2)} \\ &\simeq 1 + F_1(\cdot_2) + (1-\delta) + \beta F_{11}(\cdot_2)U_1(\cdot_2)/U_{11}(\cdot_2) > 1,\end{aligned}$$

The inequality follows from the fact that in a vicinity of the steady state $F_1(\cdot_t) + (1-\delta) \simeq 1/\beta$, $U_{11}(\cdot_t) = U_{11}(\cdot_{t+1})$, etc. This says that the impact of choosing a value for k_2 that is too large will be magnified on k_3. Employing (6.12.4) again for when $t = 2$, then gives

$$\begin{aligned}\frac{dk_4}{dk_2} &= \frac{U_{11}(\cdot_2)[dk_3/dk_2 - F_1(\cdot_2) - (1-\delta)]}{\beta[F_1(\cdot_3) + (1-\delta)]U_{11}(\cdot_3)} \\ &+ \frac{\{\beta[F_1(\cdot_3) + (1-\delta)]^2 U_{11}(\cdot_3) + \beta F_{11}(\cdot_3)U_1(\cdot_3)\} dk_3/dk_2}{\beta[F_1(\cdot_3) + (1-\delta)]U_{11}(\cdot_3)} \\ &\simeq 1 + \beta F_{11}(\cdot_2)U_1(\cdot_2)/U_{11}(\cdot_2) \\ &+ \{F_1(\cdot_3) + (1-\delta) + \beta F_{11}(\cdot_3)U_1(\cdot_3)/U_{11}(\cdot_3)\} dk_3/dk_2 \\ &> \frac{dk_3}{dk_2}.\end{aligned}$$

Hence, the error in the startup value for the difference equation will cascade into the future in an explosive manner. The situation is portrayed in Figure 6.12.1.

Remark 6.5. (Setting the value for T) Avoid the temptation to set a large value for T. Unless an *exact* solution for the model is found, the time path for the capital stock will eventually behave in an explosive manner as the above discussion highlights. So, the idea is to find an inexact time path that approximates the true time path over some reasonable length of time.

Reverse Shooting

The difference equation (6.12.1) could also be run backwards. Specifically, note that one could use (6.12.1) to solve for a value of k_t^j given values for k_{t+1}^j and k_{t+2}^j. Again, this describes one equation in one unknown. Represent the solution by

$$k_t = \overleftarrow{D}(k_{t+1}, k_{t+2}; z_{t+1}, z_t).$$

Start the system off at time T from the terminal condition $k_T = k^*$ and $z_T = z^*$. Given a value for k_{T-1} one could run the above difference

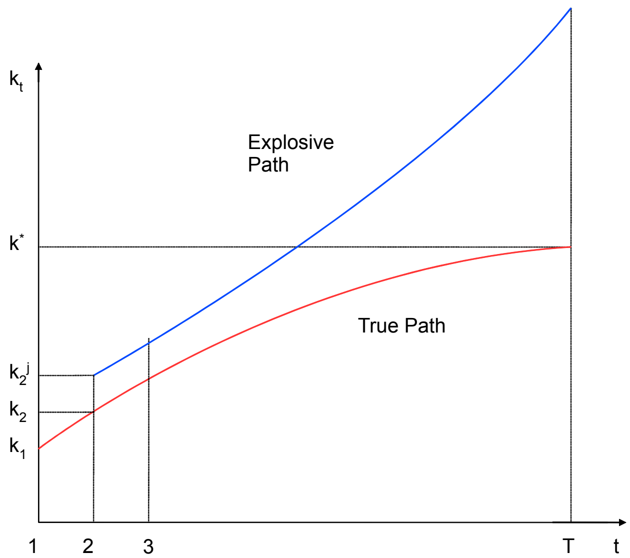


Figure 6.12.1: Multiple Shooting

equation backwards in time to get $k_{T-2}, k_{T-3}, \cdots, k_1$. This iterative scheme can be thought of as yielding a solution for k_1 as a function of k_{T-1} and k^*. Represent this function by $K_1(k_{T-1}, k^*; z_T, z_{T-1}, \cdots, z_1)$. Clearly, k_{T-1} should be chosen so that

$$K_1(k_{T-1}, k^*; z_T, z_{T-1}, \cdots, z_1) - k_1 = 0;$$

that is, when the difference equation is run backward it should go through the initial condition k_1 at time 1.

Reverse Shooting Algorithm

Find the Zero of $K_1(k_{T-1}, k^*; z_T, z_{T-1}, \cdots, z_1) - k_1 = 0$

Variables		
Predetermined	Guess	Target
$k_T = k^*$	k_{T-1}	$K_1(k_{T-1}, k^*; z_T, z_{T-1}, \cdots, z_1) = k_1$

Loop inside Function, $K_1(k_{T-1}, k^*; z_T, z_{T-1}, \cdots, z_1)$

Time, t	Variables				
	Predetermined				Computed
$T-2$	k_T		k_{T-1}		k_{T-2}
		$\swarrow$		$\swarrow$	
$T-3$	k_{T-1}		k_{T-2}		k_{T-3}
$\vdots$	$\vdots$		$\vdots$		$\vdots$
1	k_3		k_2		k_1

6.13 A Dynamic Monopoly Problem

The monopoly problem presented in Chapter 2 is now made dynamic. In each period t the monopolist faces the linear demand function

$$p_t = \alpha - \frac{\beta}{2} o_t,$$

where p_t is the period-t price of the product and o_t is the monopolist's output in this period. Demand is decreasing in price, p_t. The monopolist now produces according to the quadratic cost function

$$c_t = \frac{\gamma}{2}(o_t - \kappa o_{t-1})^2,$$

where c_t is period-t total cost and o_{t-1} is the monopolist's level of output in period $t-1$. This cost function introduces a dynamic element into the analysis. By producing more in period $t-1$ the monopolist can reduce his costs in period t. Think about this as adding learning by doing into the analysis.[10]

[10] For the learning-by-doing interpretation to make sense, assume that $o_t > \kappa o_{t-1}$. This will be the case in the setting discussed here.

The monopolist's period-t revenue is

$$p_t o_t = \alpha o_t - \frac{\beta}{2} o_t^2.$$

This implies that his period-t profits, π_t, read

$$\pi_t = \alpha o_t - \frac{\beta}{2} o_t^2 - \frac{\gamma}{2}(o_t - \kappa o_{t-1})^2.$$

Therefore, the monopolist's maximization problem is to pick his output in each period to maximize the present value of his profits. Suppose that the monopolist's discount factor is δ. The mathematical transliteration of this problem is

$$\max_{\{o_t\}_{t=1}^{\infty}} \{\sum_{t=1}^{\infty} \delta^{t-1} [\alpha o_t - \frac{\beta}{2} o_t^2 - \frac{\gamma}{2}(o_t - \kappa o_{t-1})^2]\}.$$

Note that o_t appears in exactly two periods in this optimization problem: in periods t and $t+1$. To see this, write out the objective function as

$$\cdots + \delta^{t-1}[\alpha o_t - \frac{\beta}{2} o_t^2 - \frac{\gamma}{2}(o_t - \kappa o_{t-1})^2] + \delta^t [\alpha o_{t+1} - \frac{\beta}{2} o_{t+1}^2 - \frac{\gamma}{2}(o_{t+1} - \kappa o_t)^2] + \cdots . \tag{6.13.1}$$

The first-order condition associated with this maximization problem is

$$\underbrace{\alpha - \beta o_t}_{\text{MR}} = \underbrace{\gamma(o_t - \kappa o_{t-1}) - \delta \gamma \kappa (o_{t+1} - \kappa o_t)}_{\text{MC}}, \tag{6.13.2}$$

which sets marginal revenue, MR, equal to marginal cost, MC. Observe that when the monopolist increases his output in period t he will reduce his cost in period $t+1$ (at least when $o_{t+1} > \kappa o_t$). The above first-order condition represents a linear 2nd-order difference equation in output.

6.13.1 *Steady-State Output*

Let o^* denote the steady-state level of output. It solves the equation

$$\alpha - \beta o^* = \gamma(1-\kappa)o^* - \delta\gamma\kappa(1-\kappa)o^*,$$

implying

$$o^* = \frac{\alpha}{\beta + \gamma(1-\kappa) - \delta\gamma\kappa(1-\kappa)} = \frac{\alpha}{\beta + \gamma(1-\kappa)(1-\delta\kappa)}.$$

The steady-state level of output is increasing in κ. So, adding learning in the model raises the steady-state level of output. The dynamic monopoly problem will now be solved numerically in three ways: namely, using the extended path method, multiple shooting, and the decision-rule approach. As in Chapter 2, set $\alpha = 1.0$, $\beta = 0.5$, and $\gamma = 0.5$. The new parameter, κ, governing learning by doing is selected so that $\kappa = 0.9$. The steady-state level of output is $o_T = o^* = 1.9732$.

6.13.2 *Extended Path Method*

Now rewrite the second-order difference equation connected with the monopolist's Euler equation (6.13.2) as

$$o_t = \frac{\alpha}{\beta + \gamma + \delta\gamma\kappa^2} + \frac{\gamma\kappa}{\beta + \gamma + \delta\gamma\kappa^2} o_{t-1} + \frac{\delta\gamma\kappa}{\beta + \gamma + \delta\gamma\kappa^2} o_{t+1}. \quad (6.13.3)$$

This can be expressed more compactly:

$$o_t = g + ho_{t-1} + io_{t+1},$$

where $g \equiv \alpha/(\beta + \gamma + \delta\gamma\kappa^2)$, $h \equiv (\gamma\kappa)/(\beta + \gamma + \delta\gamma\kappa^2)$, and $i \equiv (\delta\gamma\kappa)/(\beta + \gamma + \delta\gamma\kappa^2)$. Given values for o_{t-1} and o_{t+1}, this difference equation gives a solution for o_t. In any period t, the past level of output, o_{t-1}, will be known. The future value for output, o_{t+1}, is read off of a guess path, which for iteration j is denoted by $\{o_t^j\}_{t=0}^T$. Note that the guess for the final period is set so that $o_T^j = o^*$, because the output in period T is set equal to its steady-state value. The guess for the initial period is zero so that $o_0^j = 0$. So, in iteration j the above difference equation can be started off from $o_0 = 0$ to get a solution for o_1, while setting $o_2 = o_2^j$. One can then move to period 2. Here, one solves for o_2, using the previous answer for o_1, while setting $o_3 = o_3^j$. One proceeds down the path up for all $t \leq T-1$. This gives a time path $\{o_t\}_{t=0}^T$, where again by construction $o_0 = 0$ and $o_T = o^*$. Next, check whether the difference between $\{o_t\}_{t=0}^T$ and $\{o_t^j\}_{t=0}^T$ is sufficiently small. If not, set $\{o_t^{j+1}\}_{t=0}^T = \{o_t\}_{t=0}^T$ and proceed on to iteration $j+1$.

Figure 6.13.1 shows time path for output following the monopolist's introduction of a new good; i.e., $o_0 = 0$. The initial guess for output is given by a straight line between 0 and o^*. The diagram shows how the time path for output over the iterations. As can be seen, the time paths converge quickly. Additionally, it does not take many periods for the time paths to reach the steady state. Output rises over time as production costs fall due to the learning by doing. As costs fall so do prices.

The pseudo code for the extended path method is as follows.

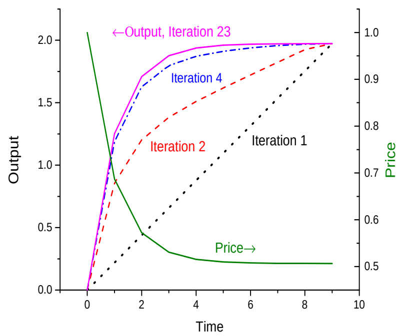


Figure 6.13.1: The solution to the dynamic monopoly problem when the extended path method is used. The initial guess for the output path is given by iteration 1. Iteration 23 shows the final solution for the output path. The solution for the final price path associated with the final output path is also shown.

Extended-path method, pseudo code

1. Input values for α, β, and γ and calculate values for g, h, and i. Declare g, h, and i as global variables. Set the terminal time period T, also a global variable. Do not set T too large.

2. Set the initial guess for the time path of output. Output in the first period and terminal periods are 0 and o^*.

3. Write a function for the time path that takes a guess for the output time path as input and returns a revised guess for the time path as output. The function inputs g, h, i, and T as global variables. The function iterates on the difference equation (6.13.3) starting at period 2 and terminating in period $T - 1$. This is done using a for loop. In each period t the difference equation takes o_{t-1} from the solution for the previous period and inserts o_{t+1} from the guess path.

4. Write a while loop that iterates on guess paths. Specifically, the while loop takes guess path and computes a revised guess using the above function. The while loop should specify a maximum number of iterations to prevent an infinite loop from occurring.

5. From the convergent time path for output, compute other variables of interest such as a time path for prices.

6. Plot the results, in particular the time paths for output and prices.

6.13.3 *Multiple Shooting*

To setup the problem for multiple shooting, the second-order difference equation for output, associated with the Euler equation (6.13.2) arising from monopolist's problem is rewritten as

$$o_{t+1} = -\frac{\alpha}{\delta\gamma\kappa} + \frac{\beta + \gamma + \delta\gamma\kappa^2}{\delta\gamma\kappa} o_t - \frac{\gamma\kappa}{\delta\gamma\kappa} o_{t-1}, \tag{6.13.4}$$

or

$$o_{t+1} = d + e o_t + f o_{t-1},$$

with $d \equiv -\alpha/(\delta\gamma\kappa)$, $e \equiv (\beta + \gamma + \delta\gamma\kappa^2)/(\delta\gamma\kappa)$, and $f \equiv -\gamma\kappa/(\delta\gamma\kappa)$. This second-order difference equation can be solved using multiple shooting. Given two starting conditions, o_0 and o_1, the above equation can be iterated forward in time to get o_2, o_3, o_4,$\cdots$. The monopolist only starts producing in period 1 so $o_0 = 0$. The idea underlying multiple shooting is to select o_1 so that $\lim_{t\to\infty} o_t = o^*$; i.e., so that as the end of time approaches the time path of output converges to the steady-state level of output. This is operationalized by running the model over some finite time horizon. In particular, say T periods. Output in the first period, o_1, is then chosen, using a nonlinear equation solver, so that output in the last period, T, is equal to the steady-state level of output, or $o_T = o^*$. It is a good idea not to set T too large because the time path for output will often have an explosive behavior when trying out guesses for o_1.

Dynamic Monopoly, Multiple Shooting-pseudo code

The pseudo code for solving the dynamic monopoly problem using multiple shooting is below. The convergent time path for output is once again shown in 6.13.1.

1. Input values for α,β, and γ and calculate values for d, e, and f. Declare d, e, and f as global variables. Also set the initial level of output, $o_1 = 0$, the steady-state level of output, o^*, and the time horizon, T, as global variables. Do not set T too large.

2. Write a function for the time path that takes a guess for output in the second period, o_2, and returns the deviation of terminal level of output, o_T, from the steady state, o^*. The functions specifies d, e, f, o_1, o^*, and T to be global variables The function iterates on the difference equation (6.13.4) starting at period 3 and terminating in period T. For period 3 the function uses the starting conditions o_1 and o_2. Save the time path for output as a global variable.

3. Use a nonlinear equation solver to compute a value for o_2 while calling the above function.

4. From the convergent time path for output, compute other variables of interest such as a time path for prices.

5. Plot the results, in particular the time paths for output and prices.

6.13.4 *Decision-Rule Approach*

The decision rule approach is discussed now. Conjecture that the monopolist's decision rule has the following linear form:

$$o_t = \eta + \psi o_{t-1}. \tag{6.13.5}$$

In a steady state,

$$o^* = \frac{\eta}{1-\psi}.$$

Hence, the constant η must solve

$$\eta = \frac{(1-\psi)\alpha}{\beta + \gamma(1-\kappa) - \delta\gamma\kappa(1-\kappa)}. \tag{6.13.6}$$

If this is the case, then one can rewrite the first-order condition as

$$\alpha - \beta o_t = \gamma(o_t - \kappa o_{t-1}) - \delta\gamma\kappa(\eta + \psi o_t - \kappa o_t).$$

Therefore,

$$(\gamma + \beta - \delta\gamma\kappa\psi + \delta\gamma\kappa^2)o_t = \alpha + \delta\gamma\kappa\eta + \gamma\kappa o_{t-1},$$

so that

$$o_t = \frac{\alpha + \delta\gamma\kappa\eta}{\gamma + \beta - \delta\gamma\kappa\psi + \delta\gamma\kappa^2} + \frac{\gamma\kappa}{\gamma + \beta - \delta\gamma\kappa\psi + \delta\gamma\kappa^2} o_{t-1}.$$

This implies that

$$\psi = \frac{\gamma\kappa}{\gamma + \beta - \delta\gamma\kappa\psi + \delta\gamma\kappa^2}.$$

Therefore, the solution for ψ solves the quadratic equation

$$-\delta\gamma\kappa\psi^2 + (\gamma + \beta + \delta\gamma\kappa^2)\psi - \gamma\kappa = 0.$$

Since this is a quadratic equation there will be two roots. Observe that the lefthand side of the above equation is negative when $\psi = 0$ and that its derivative is positive at this point. The expression is positive when $\psi = 1$ (since the steady-state level of output must be positive). The lefthand side eventually becomes negative as ψ becomes large. It is easy to solve this equation for ψ on the computer. Given the solution for ψ one can recover the solution for η using (6.13.6). The time path for output is then obtained by iterating on (6.13.5) starting from $o_0 = 0$.

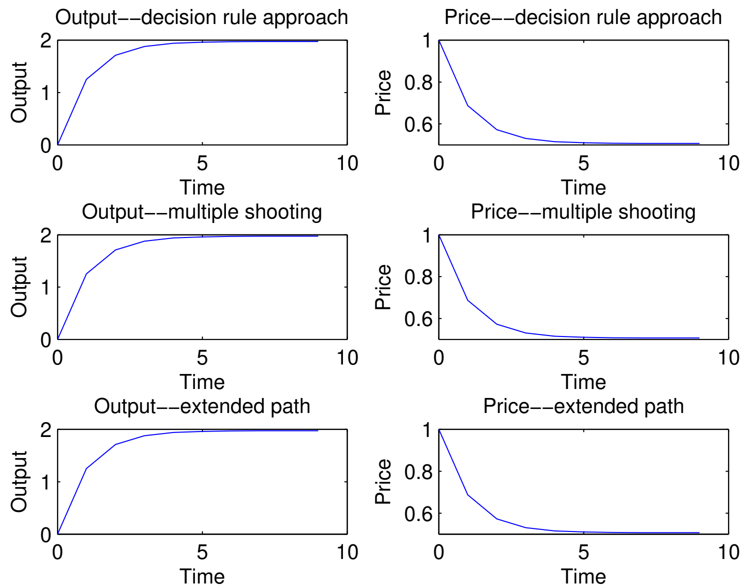


Figure 6.13.2: Output for the dynamic monopoly model. Multiple shooting, the extended path method, and the decision-rule approach all give the same solution for the time path of output and prices.

6.13.5 *Results*

The is the output for the dynamic monopoly model for the three methods is shown in Figure 6.13.2. The three techniques give the identical solution for the time path of output and prices.

6.14 *Problem*

Consider the problem of a consumer/worker whose lifetime utility function is given

$$\sum_{t=0}^{\infty} \beta^t U\Big(c_t - G(h_t)\Big), \text{ with } 0 < \beta < 1,$$

where c_t is his period-t consumption and h_t is his hours worked in period t. Let,

$$U\Big(c - G(h)\Big) = \ln\Big(c - G(h)\Big).$$

$$G(h) = \frac{h^{1+\theta}}{1+\theta},$$

and

$$\beta = \frac{1}{1+\rho},$$

where ρ is the person's rate of time preference. In period t the worker owns capital, k_t, which he rents out to a firm.

There is a firm in the economy that produces output in period t, o_t, according to the production function

$$o_t = F(k_t, h_t) = k_t^{\alpha} h_t^{1-\alpha}, \text{ where } 0 < \alpha < 1.$$

The firm hires capital, k_t, and labor, h_t, from the consumer/worker. The firm pays w_t for each unit of labor that it hires and r_t for each unit of capital that it rents.

Last, there is government in the economy that taxes the worker's labor income at rate τ_h. It also taxes his rental income from capital at the rate τ_k. The worker's capital depreciates at rate δ over time. The tax code allows the consumer/worker to deduct this depreciation from his rental income when calculating his taxable income. The government rebates back transfer payments in the form of lump-sum transfer payments, λ_t.

1. Set up and solve the representative agent's and firm's maximization problems. Characterize the economy's general equilibrium.

2. Derive a formula for the economy's steady-state capital stock, k^*, and labor supply, h^*. How do taxes affect these quantities?

3. Set $\alpha = 0.3$, $\beta = 1/(1.04)$, $\delta = 0.08$, $\theta = 0.6$, $\tau_h = 0.25$ and $\tau_k = 0.15$. What are the economy's steady-state levels of capital, hours worked and GDP?

4. President Noah Nutting is proposing to increase the tax rate on capital and labor $\tau_h = 0.30$ and $\tau_k = 0.30$. What will the new steady state look like? Compute the economy's time path for capital, consumption, hours worked, and GDP from the old to new tax regime. Use the *extended path* method.

5. Congresswoman T. H. Ink suggests that it in order to raise the extra 5% in tax revenue that the President wants for his "Middle Class America" program it would be better to eliminate the capital income tax and raise the labor income tax to $\tau_h = 0.315$. She says that the President is on the verge of "killing the goose that lays the golden egg" by dissuading investment in the economy. The president retorts that this is a "reversion to the old way of thinking." Who is right? Why

7 Malthus to Solow

"In October 1838, that is, fifteen months after I had begun my systematic inquiry, I happened to read for amusement Malthus on Population, and being well prepared to appreciate the struggle for existence which everywhere goes on from long- continued observation of the habits of animals and plants, it at once struck me that under these circumstances favourable variations would tend to be preserved, and unfavourable ones to be destroyed. The results of this would be the formation of a new species. Here, then I had at last got a theory by which to work."
Charles Darwin (1876)

7.1 Introduction

The goal here is to model the transition from a world where living conditions were stagnant over a long period of time to a world with rising living standard.[1] The analysis presumes the existence of two technologies: Malthus and Solow. The preindustrial era uses a land-intensive technology. This constant-returns-to-scale technology also employs capital and labor. Productivity for this technology grows at a very slow rate. Land is in fixed supply. The necessity to use land places a drag on growth. This technology is dubbed the Malthus technology. The modern era uses a constant-returns-to-scale technology employing just capital and labor. Productivity for this technology grows at a faster clip than the preindustrial one, although its productivity starts off from a very low level. This technology is labeled the Solow technology.- Both technologies are *always* available. At low levels of development it pays only to use the Malthus technology. As the economy develops it becomes profitable also to use the Solow technology. The Malthus technology fades away asymptotically.

[1] This section is based on Hansen and Prescott (2002).

Thomas R. Malthus (1766-1834) was an English cleric and economist. He write the famous book *An Essay on the Principle of Population*, which postulated that the size of a population is limited by the productive capacity of land. Charles Darwin credits Malthus's work as being instrumental in forming his theory of evolution.

Robert M. Solow (1923-) is an American economist who is best known for his work on economic growth. He broke down the sources of economic growth into changes in the labor supply, increases in the capital stock, and technological progress. He won the Nobel Prize in 1987.

7.2 A Graphical Exposition of Malthusian Theory

Malthusian theory states the size of the population will be regulated by the productive capacity of the economy. Figure 7.2.1 portrays the situation. Start with the lower panel. Income per worker, y, is negatively related to the number of workers, n, as the curve in the lower panel of the diagram portrays. This occurs because land is fixed in supply. Turn to the upper panel. Fertility is increasing in income because parents can better support larger families. Likewise, the mortality rate declines in income, since the diseases related to poverty fall. The per-worker level of income associated with a stable population size, n^*, is given by y^*. If income per worker was at some higher level, say y', then population size would increase. This would occur since fertility would exceed mortality. This expansion in population would lead to a decline in income per worker until it converges to y^*.

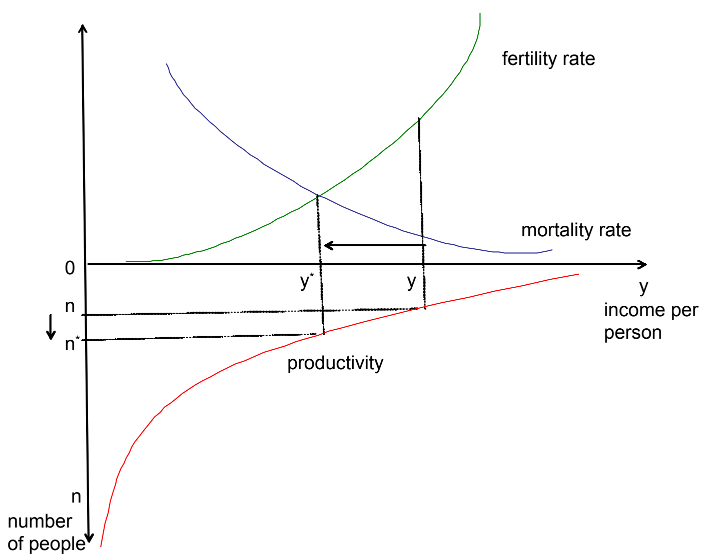


Figure 7.2.1: Malthusian equilibrium

7.3 *Facts*

7.3.1 *England, 1275-1800–Malthus Era*

Real wages were roughly constant for a long period of time in preindustrial England. When population fell, in the Black Death, real wages rose. This is in accord with Malthusian theory. Here wages adjust to limit the size of the population–see Figure 7.3.1. Malthusian theory predicts that population and land rents will rise and fall together. They did over this period–see Figure 7.3.2.

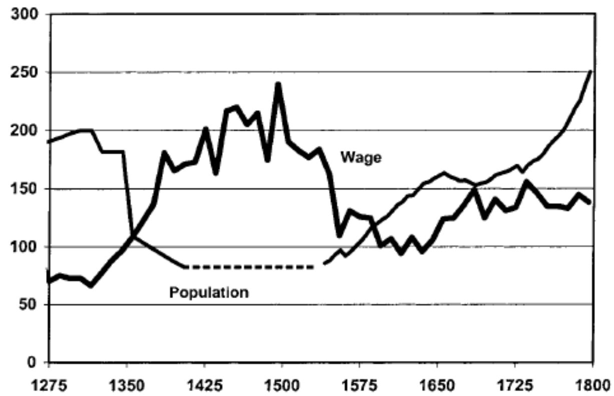


Figure 7.3.1: Population and real wages: England, 1279-1800. *Source*: Hansen and Prescott (2002).

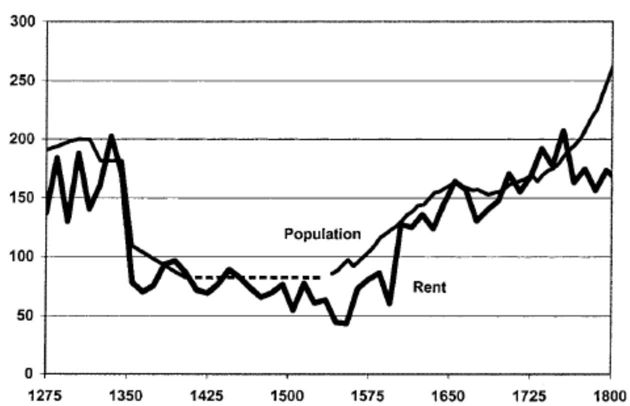


Figure 7.3.2: Population and real land rents: England, 1279-1800. *Source*: Hansen and Prescott (2002).

7.3.2 England 1800-1989–Solow Era

By contrast, in the industrial era population growth did not lead to falling real wages as Malthusian theory predicts–see Figure 7.3.3. It's hard to see a relationship between population growth and labor productivity. The Solow model doesn't predict one–Figure 7.3.4. The value of farmland to GDP fell–see Figure 7.3.5.

7.4 The Model

The vehicle for analysis is a two-period overlapping generations model. At any point of time, there are two generations of adults, young and old. People only work while young. They supply one unit of labor which earns a wage. Old people are retired. They live off of the savings they undertook when young. The young can save in the form of capital or land. Capital depreciates fully across periods. Capital is reproducible. Land lasts forever. It is fixed in supply at one.

There are two technologies, a primitive one and an advanced one. Either or both can be operated in a period. The primitive technology is labeled as the Malthus technology Here output, y_m, is produced using capital, k_m, labor, n_m, and land, l_m, according to

$$y_m = a_m k_m^{\phi} n_m^{\mu} l_m^{1-\phi-\mu}.$$

The key factor in the Malthus technology is its use of land, l_m. The advanced technology is dubbed the Solow technology. Under the Solow technology production is governed by

$$y_s = a_s k_s^{\theta} n_s^{1-\theta},$$

where y_s is the level of output produced by the Solow technology, and k_s, and n_s are the inputs of capital, and labor. The Solow technology does not use land. It is assumed that a_s grows at a faster rate over time than a_m. Let the gross growth rate of technological progress in the Solow sector be represented by $\gamma_s \equiv a_s'/a_s$ and that for the Malthus by $\gamma_m \equiv a_m'/a_m$. (Here the prime or $'$ denotes the value of the variable next period.) Since the Solow technology has a faster rate of technological progress, $\gamma_s > \gamma_m$. So, on this accounts, one would expect growth to be faster with the Solow technology. The fact that land is not a reproducible factor slows down growth further in the Malthus technology.

Last to complete the setup, the economy's resource constraint is presented

$$\mathbf{c} + \mathbf{k}' = y_m + y_s.$$

Here $\mathbf{c}$ is *aggregate* consumption and $\mathbf{k}'$ is *aggregate* investment. Aggregate consumption is the sum of consumption over the old and young

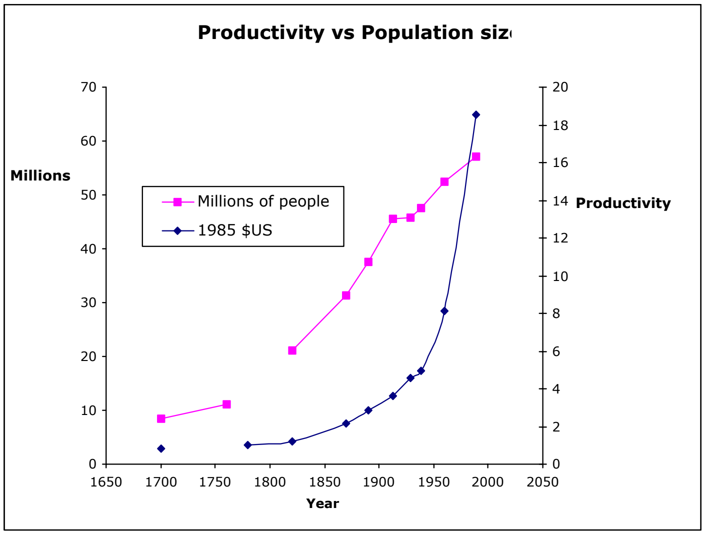


Figure 7.3.3: The relationship between population and productivity, 1700-1989

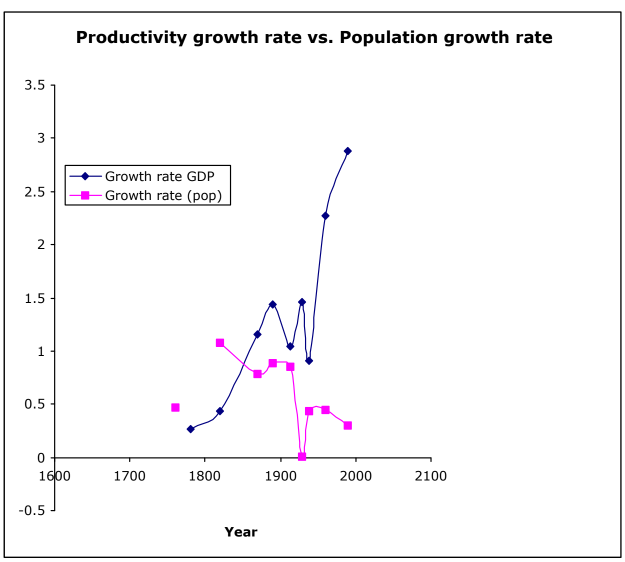


Figure 7.3.4: The relationship between population growth and productivity growth

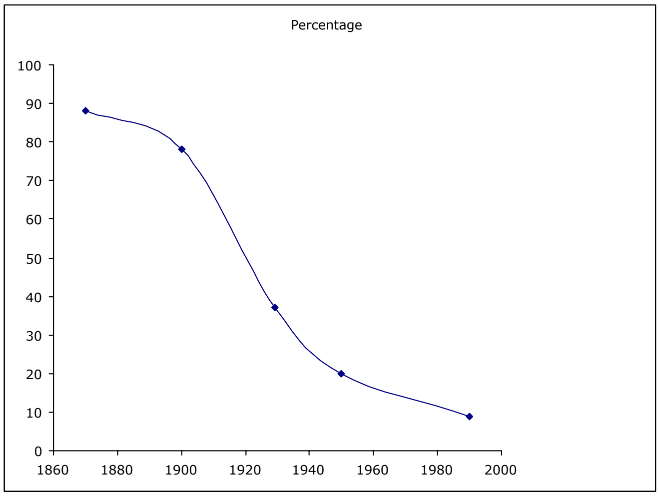


Figure 7.3.5: Value of farm land relative to GDP, U.S., 1870-1990

generations, which will be of different sizes. Aggregate investment is just done by the young. Therefore, aggregate demand, $\mathbf{c} + \mathbf{k}'$, must equal aggregate supply, $y_m + y_s$.

7.4.1 *Firms' Problems*

Firms hire capital, labor and land to maximize profits. They solve the following maximization problems:

$$\max_{k_m, n_m, l_m} \{a_m k_m^{\phi} n_m^{\mu} l_m^{1-\phi-\mu} - w n_m - r_k k_m - r_l l_m\}, \tag{7.4.1}$$

and

$$\max_{k_s, n_s} \{a_s k_s^{\theta} n_s^{1-\theta} - w n_s - r_k k_s\}, \tag{7.4.2}$$

where w is the wage rate and r_k and r_l are the rental rates on capital and land.

7.4.2 *Household's Problem*

Each household solves the following maximization problem

$$\max_{c_1, c_2', k', l'} \{\ln c_1 + \beta \ln c_2'\}, \tag{7.4.3}$$

subject to their first- and second-period budget constraints

$$c_1 + k' + q l' = w,$$

and

$$c_2' = r_k' k' + (r_l' + q') l'.$$

Here q is the current price for a unit of land. The first-period budget constraint states that when young consumption, c_1, and savings in capital and land, $k' + ql'$, equals labor income, w. The second-period budget constraint says that when old the person's consumption, c_2', will be limited by the income he earns from his ownership of capital and land, $r_k' k' + (r_l' + q')l'$. When old the person will sell his land (to a young person in the next generation) at the unit price q'.

7.4.3 Demographics

Population growth is simply given by

$$n' = G(c_1)n. \tag{7.4.4}$$

The function $G(c_1)$ is constructed to match the demographic transition. The demographic transition refers to the fact that fertility has had $\cap$ shape over time (or equivalently has risen and then fallen with living standards).

7.4.4 Equilibrium

Let n represent the size of the today's young population and n_{-1} the size of the old population. An equilibrium for the above economy must satisfy the following conditions.

1. Firms maximize profits or solve problems (7.4.1) and (7.4.2).
2. The households maximize utility or solve problem (7.4.3).
3. All markets clear implying
   (a) Physical Capital
   $$k_m + k_s = n_{-1}k,$$
   (b) Labor
   $$n_m + n_s = n,$$
   (c) Land
   $$n_{-1}l = 1,$$
   (d) Goods
   $$nc_1 + n_{-1}c_2 + nk' = y_m + y_s,$$
   where aggregate consumption and investment are given by $\mathbf{c} = nc_1 + n_{-1}c_2$ and $\mathbf{k}' = nk'$.

7.5 Malthus versus Solow

The cost function for the Solow sector is

$$C_s(w, r_k, y_s) = \min_{k_s, n_s}\{r_k k_s + w n_s : y_s = a_s k_s^{\theta} n_s^{1-\theta}\} = a_s^{-1}\theta^{-\theta}(1-\theta)^{-(1-\theta)} r_k^{\theta} w^{1-\theta} y_s.$$

This cost function has the standard properties. Cost is increasing and concave in both prices, r_k and w, separately. It is homogenous of degree one in both prices together. That is, if both prices are increased by a factor λ, then costs will rise by this factor too. Cost also rises with the level of output, y_s, produced. Here marginal cost is

$$a_s^{-1}\theta^{-\theta}(1-\theta)^{-(1-\theta)}r_k^{\theta}w^{1-\theta}.$$

Hence, marginal cost is constant.

The cost function for the Malthus sector (holding land fixed at unity) is

$$\begin{aligned} C_m(w, r_k, y_m) &= \min_{k_m, n_m}\{r_k k_m + w n_m : y_m = a_m k_m^{\phi} n_m^{\mu} l_m^{1-\phi-\mu} \text{ and } l_m = 1\} \\ &= a_m^{-1/(\phi+\mu)}[(\frac{\phi}{\mu})^{\mu/(\phi+\mu)} + (\frac{\phi}{\mu})^{-\phi/(\phi+\mu)}]r_k^{\phi/(\phi+\mu)}w^{\mu/(\phi+\mu)}y_m^{1/(\phi+\mu)}, \end{aligned}$$

so that marginal cost will be

$$\frac{1}{(\phi+\mu)}a_m^{-1/(\phi+\mu)}[(\frac{\phi}{\mu})^{\mu/(\phi+\mu)} + (\frac{\phi}{\mu})^{-\phi/(\phi+\mu)}]r_k^{\phi/(\phi+\mu)}w^{\mu/(\phi+\mu)}y_m^{1/(\phi+\mu)-1}.$$

Here, marginal cost is increasing and convex. Observe that marginal cost goes to zero as output goes to zero.

The Solow sector will not operate when

$$\begin{aligned} a_s^{-1}\theta^{-\theta}(1-\theta)^{-(1-\theta)}r_k^{\theta}w^{1-\theta} > &\frac{1}{(\phi+\mu)}a_m^{-1/(\phi+\mu)}[(\frac{\phi}{\mu})^{\mu/(\phi+\mu)} + (\frac{\phi}{\mu})^{-\phi/(\phi+\mu)}] \\ &\times r_k^{\phi/(\phi+\mu)}w^{\mu/(\phi+\mu)}y_m^{1/(\phi+\mu)-1}. \end{aligned}$$

That is, the Solow sector will not operate at any aggregate output levels, y_s, where the Solow sector has higher marginal cost. Both sectors will operate only when

$$\begin{aligned} a_s^{-1}\theta^{-\theta}(1-\theta)^{-(1-\theta)}r_k^{\theta}w^{1-\theta} = &\frac{1}{(\phi+\mu)}a_m^{-1/(\phi+\mu)}[(\frac{\phi}{\mu})^{\mu/(\phi+\mu)} + (\frac{\phi}{\mu})^{-\phi/(\phi+\mu)}] \\ &\times r_k^{\phi/(\phi+\mu)}w^{\mu/(\phi+\mu)}y_m^{1/(\phi+\mu)-1}. \end{aligned}$$

The Malthus sector will always operate since, as was mentioned, its marginal cost goes to zero as output goes to zero. Figure 7.5.1 shows the adoption point, at a given set of factor prices.

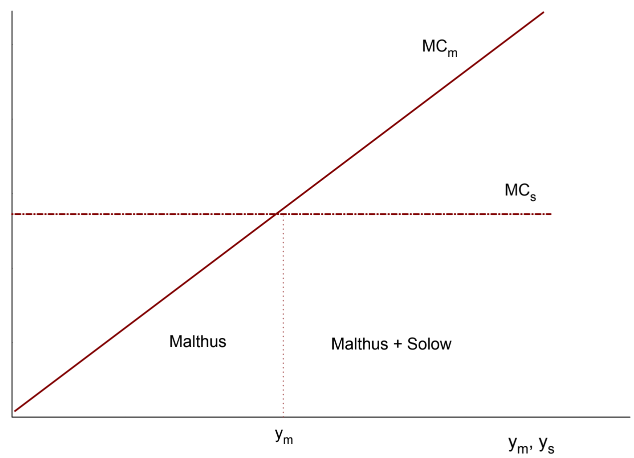


Figure 7.5.1: The Solow Adoption Point

Lemma 7.1. *The Solow technology is not used if*

$$a_s < (\frac{r_k}{\theta})^{\theta}(\frac{w}{1-\theta})^{1-\theta}.$$

Proof. (Gokce Uysal, Richard Suen and Vikram Manjunath) Imagine one is in a world where only the Malthus technology is used and the factor prices are given by r_k and w. It will be shown that when the above condition holds it is not optimal to use the Solow technology. First, it will be shown that profits will be negative if the above condition holds. In the Solow sector $\theta = r_k k_s / y_s$ and $1 - \theta = w n_s / y_s$. Therefore, profits can be written as

$$a_s(\frac{\theta y_s}{r_k})^{\theta}(\frac{(1-\theta)y_s}{w})^{1-\theta} - \frac{\theta y_s}{k_s}k_s - \frac{(1-\theta)y_s}{n_s}n_s.$$

For any $y_s > 0$ profits will be negative whenever

$$a_s < (\frac{r_k}{\theta})^{\theta}(\frac{w}{1-\theta})^{1-\theta}.$$

An alternative proof can be constructed that demonstrates that the first-order conditions for a firm using the Solow technology will be violated if the above inequality holds. Suppose that the statement in the lemma holds and that the Solow sector operates. The first-order conditions to (7.4.2) imply that

$$\frac{n_s}{k_s} = (\frac{r_k}{\theta})/(\frac{w}{1-\theta}).$$

Therefore, from the first-order condition for capital

$$a_s\theta(\frac{k_s}{n_s})^{\theta-1} = a_s\theta(\frac{r_k}{\theta})^{1-\theta}(\frac{w}{1-\theta})^{\theta-1} = r_k.$$

This implies that

$$a_s = (\frac{r_k}{\theta})^{\theta}(\frac{w}{1-\theta})^{1-\theta},$$

a contradiction.

Yet, a third proof follows from the fact that the marginal cost of producing in the Solow sector cannot exceed the price of output, which is one, so that $a_s^{-1}\theta^{-\theta}(1-\theta)^{-(1-\theta)}r_k^{\theta}w^{1-\theta} \leq 1$. □

7.6 *The Malthusian Steady State*

Before proceeding on further, the Malthusian steady state will be formulated. First, it is easy to define the rate of population growth in the Malthusian steady state. In a Malthus-*only* economy population grows at the same rate as output, γ_{y_m}. Now,

$$\gamma_{y_m} \equiv \frac{y'_m}{y_m} = \frac{a'_m k'^{\phi}_m n'^{\mu}_m l'^{1-\phi-\mu}_m}{a_m k^{\phi}_m n^{\mu}_m l^{1-\phi-\mu}_m}.$$

Recall that $a'_m/a_m = \gamma_m$. Land is fixed in supply so $l'_m = l_m$. Since population is growing at the same rate as output $n'_m/n_m = \gamma_{y_m}$. Last, capital will grow at the same rate as output, too, so that $k'_m/k_m = \gamma_{y_m}$. Using these facts in the above equation leads to the result $\gamma_{y_m} = \gamma_m^{1/(1-\mu-\phi)}$.

Second, the level of consumption for young adults is now immediate. Since population is growing at the same rate as output, it happens that per-capita output, and consumption for the young, c_1, are constant. The level of c_1 can be determined by the steady-state for the Malthus-only equilibrium. To characterize the Malthus-only steady state, note from (7.4.4) that $n_{t+1} = G(c_{1,t})n_t$ so

$$c_1 = G^{-1}(\gamma_{y_m}).$$

Third, the level of wages, w, can now be computed. From the consumer's problem it can be calculated that

$$c_1 = w/(1+\beta).$$

Therefore, in a Malthusian steady state

$$w = (1+\beta)c_1 = (1+\beta)G^{-1}(\gamma_{y_m}). \tag{7.6.1}$$

The steady-state wage rate has been pinned down.

Last, the rental prices for capital and land, r_k and r_l, and the price of land, q, can be determined. In a steady state the price of land will be given by $q = (r_l + q)\gamma_{y_m}/i$, where i is the gross interest rate–note that $r'_l = \gamma_{y_m} r_l$ and $q' = \gamma_{y_m} q$. Therefore,

$$q = r_l(\gamma_{y_m}/i)/(1-\gamma_{y_m}/i); \tag{7.6.2}$$

that is, the price of land is the discounted value of the (growth-adjusted) rents that it will earn. It is immediate that $r_k = i$ since the gross return on capital must equal the gross interest rate.

7.6.1 *Savings Equals Investment*

So, what determines the interest rate i? In a steady state aggregate saving by the young, $n(w - c_1)$, must equal aggregate investment in capital and land, $\gamma_{y_m} k_m + q$. (Here, $k'_m = \gamma_{y_m} k_m = nk'$.) Therefore,

$$n(w - c_1) = n\beta/(1+\beta)w = \gamma_{y_m} k_m + q,$$

or

$$(w - c_1) = \beta/(1+\beta)w = \gamma_{y_m} k_m/n + q/n. \tag{7.6.3}$$

Think about (7.6.3) as determining the interest rate i. Observe that the lefthand side is a constant, since w is just a function of γ_{y_m}. It will now be shown that k_m/n and q/n on the righthand side can be expressed as functions of i. Given this, equation (7.6.3) determines i. [Condition (7.6.3) is actually the same as the goods market-clearing condition. Try to convince yourself that this is true.] Thus, in a steady state for an overlapping generations model the interest rate, i, is not pinned down in simple fashion by the discount factor, as in the representative agent model.

By taking the ratio of the first-order conditions for labor and capital in the firm's optimization problem for the Malthus sector, it can be shown that

$$\frac{k_m}{n_m} = (w/i)(\phi/\mu).$$

In a Malthus-only equilibrium n_m is equal to the size of the young population, n. Thus,

$$k_m/n = (w/i)(\phi/\mu).$$

Therefore, k_m/n is a function of i, as claimed. The first-order conditions for land and labor in the Malthus sector imply

$$r_l = \frac{1-\phi-\mu}{\mu} wn.$$

This implies that q/n will be a function of i from (7.6.2). Hence, (7.6.3) determines $i = r_k$.

Next, the equilibrium size of the population will be uncovered. Using the formula for the marginal product of labor in the Malthus sector

$$w = \mu a_m k_m^{\phi} n^{\mu-1} = \mu a_m [n(w/i)(\phi/\mu)]^{\phi} n^{\mu-1}. \tag{7.6.4}$$

It's clear from (7.6.4) that n, or the size of the young population, can be written as a simple function of i–recall that w is pinned down by

(7.6.1). Specifically,

$$n = (\frac{1}{a_m})^{1/(\phi+\mu-1)} \mu^{(\phi-1)/(\phi+\mu-1)} w^{(1-\phi)/(\phi+\mu-1)} (\frac{i}{\phi})^{\phi/(\phi+\mu-1)}.$$

7.7 Calibration

7.7.1 Demographics–in accord with Lucas (1998)

$$G(c_1) = \begin{cases} \gamma_m^{1/(1-\mu-\phi)}(2 - \frac{c_1}{c_{1m}}) + 2(\frac{c_1}{c_{1m}} - 1), & \text{for } c_1 < 2c_{1m}, \text{ (rising segment)} \\ 2 - \frac{c_1 - 2c_{1m}}{16c_{1m}}, & 2c_{1m} \le c_1 \le 18c_{1m}, \text{ (falling segment)} \\ 1, & \text{for } c_1 > 18c_{1m} \text{ (flat segment).} \end{cases}$$

This function is plotted in Figure 7.7.1 below. The figure has three interesting features. First, at the Malthusian level of living standards, $c_1 = c_{1m}$, population grows at the same pace as output, $\gamma_m^{1/(1-\mu-\phi)}$, where γ_m is the rate of growth in a_m. Second, as living standards double from the Malthusian level the population growth rate rises to a point where it doubles every thirty-five (the period length) years. Third, from this point to the point where living standards are 18 times the Malthusian one ($2c_{1m} \le c_1 \le 18c_{1m}$) the rate of population growth declines until a stationary level is attained.

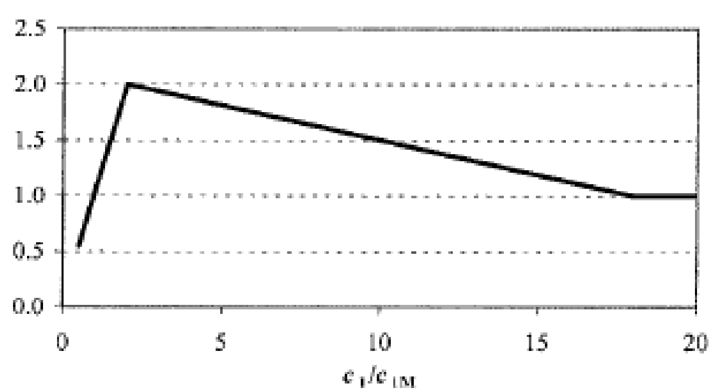


Figure 7.7.1: Demographics. *Source*: Hansen and Prescott (2002).

7.7.2 Parameter Values

The parameter values chosen are shown in the table below. Technological progress is faster in the Solow era than in the Malthus one. They are picked to mimic growth in each era. Capital's share of income is much less in the Malthus economy relative to the Solow one. This is because land is also used in the Malthus economy. The need to use land in the Malthus era slows down growth, as will be discussed below. Another point to note is that the discount factor is set to one. Still, the model gives reasonable values for the interest rate in the Malthus and Solow eras. This is because the interest rate is not pinned down by the discount factor, alone, in an overlapping generations model.

Remark 7.1. If *both* land and labor were fixed, then in an economy using only the Malthus technology output would grow at rate $\gamma_m^{1/(1-\phi)}$, while in a Solow one the growth rate would be $\gamma_s^{1/(1-\theta)}$. Thus, when $\gamma_m \geq \gamma_s$ the Solow economy grows faster because $\theta > \phi$; i.e., the reproducible factor, capital, has a larger share in the Solow economy.

Parameter	Value	Comment
γ_m	1.032	Growth in Malthus Era–period 35 yrs (0.1% a year.) Consistent with pop. doubling every 230 yrs
γ_s	1.518	Postwar GDP Growth (1.2% a year)
ϕ	0.1	Capital's Share of Income, Malthus
μ	0.6	Labor's share, both technologies
θ	0.4	Capital's Share, Solow
β	1.0	Discount factor– annual return of 2% in Malthus era, 4-4.5% in Solow era.

7.8 Results

As the economy develops, the share of inputs devoted to the Malthus sector declines over time. This is shown in Figure 7.8.1. Figure 7.8.2 shows how wages rise and the population grows as the economy moves to the Solow epoch. Last, land isn't used by the Solow technology. So, its value declines over time as the Malthus sector dies out. This is shown in Figure 7.8.3.

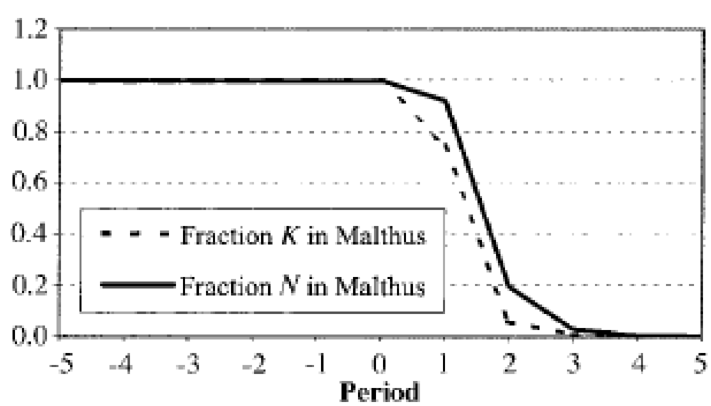


Figure 7.8.1: The vanishing of the Malthusian sector. *Source*: Hansen and Prescott (2002).

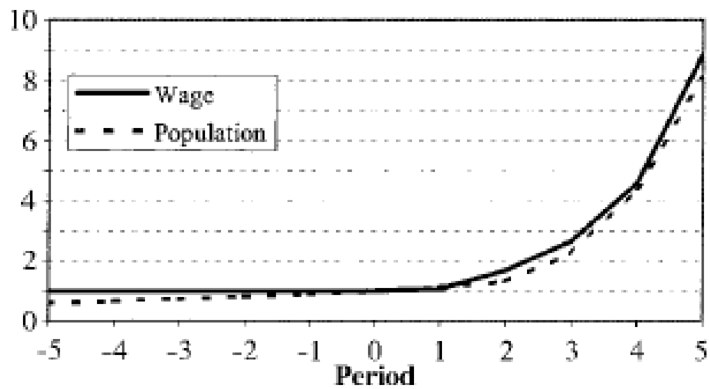


Figure 7.8.2: The Solow era. *Source*: Hansen and Prescott (2002).

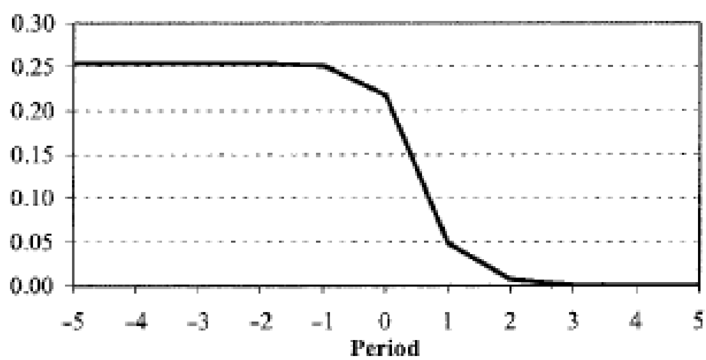


Figure 7.8.3: The declining value of land. *Source*: Hansen and Prescott (2002).

8 Numerical Approximations

8.1 Introduction

Solving macroeconomic models, especially stochastic ones, on the computer often involves employing numerical approximation for things such as derivatives, functions, and probability distributions. Some of these numerical approximations are discussed here. The discussion starts off with numerical differentiation. This is often used to log-linearize dynamic stochastic models, as will be discussed in Chapter 9. The concept of numerical differentiation is illustrated by examining the impact of technological progress in contraception on non-marital births by young women. An enigma is the U.S. data is that as contraception improved non-marital births jumped up. By comparing numerical derivatives at two different points in time, this fact can be rationalized. Computers can also differentiate symbolically analytical expressions. This is demonstated using a simple consumption/leisure choice problem. The chapter then moves on to the topic of numerical integration. This is illustrated by an example that calculates the consumer surplus for computers, which is the roughly trianglar shaped area trapped below the demand curve for computers and above the price line for computers.

Numerical integration is then followed by the topic of random number generation. Random numbers are used in Monte Carlo simulation of business cycle models, again discussed in Chapter 9. The concept of random number generation is illustrated using an example from Eugen Slutsky (1937)'s classic work on "Random Causes as the Source of Cyclic Processes." Next, the concept of Monte Carlo integration using random number generation is discussed. This concept is illustrated by calculating the welfare cost of business cycles à la Robert E. Lucas, Jr.

The chapter also introduces the idea of a Markov chain. As an example a Markov chain for the evolution of employment/unemployment is presented. When calibrated this Markov chain can be used to backout the job finding and separation rates. Mehra and Prescott (1985)'s well-known study on the equity premium is also used to illustrate the notion of a Markov chain. It is shown how an AR1 process can be approximated by a Markov chain. Last, the topic of approximating a function is discussed; e.g., some policy function or value function as in Chapter 9. Three methods are discussed: piecewise linear interpolation, cubic spline interpolation, and radial basis function interpolation. Cubic spline interpolation is very flexible. To show this, a facsimile of an artist's sketch of a face is generated using cubic splines. The Hodrick-Prescott filter, which is based on cubic splines, is also presented.

8.2 Numerical Differentiation

8.2.1 The Standard Method

Computing analytically the derivatives of a function, say $F(x)$, can take some effort. Often it is simpler to approximate these derivatives numerically. The derivative of the function $F(x)$ at the point x can be computed numerically by using the formula

$$F_1(x) = \frac{F(x+h) - F(x-h)}{2h},$$

where h is some small number. Computing numerical derivatives is a bit trickier than this formula suggests. Mathematically speaking, one would like to make h as small as possible. But, note that the difference between $F(x+h)$ and $F(x-h)$ will be rounding error if h is made too small. This occurs because the numbers $F(x+h)$ and $F(x-h)$ will be computed with small errors at the nth decimal where n is some integer, say 10. If h is made too small the difference will just be this error. Dividing this through by a very small h then blows this up. Hence, there is a trade-off. Making h small improves mathematical precision but increases numerical error. So, how small should h be? A good lower bound on a PC would be about 1.0e-5.

To see the problem formally, take a first-order Taylor expansion of the function F around the point x. (Chapter A reviews the concept of a first-order Taylor expansion.) One gets

$$F(x+h) = F(x) + F_1(x)h + F_{11}(\zeta)h^2/2 \text{ (for } x \leq \zeta \leq x+h).$$

In a similar fashion, one can write

$$F(x-h) = F(x) - F_1(x)h + F_{11}(\xi)h^2/2 \text{ (for } x-h \leq \xi \leq x).$$

Subtracting the second equation from the first, while applying the intermediate value theorem, gives,

$$F_1(x) = \frac{F(x+h) - F(x-h)}{2h} + \frac{[F_{11}(\xi) - F_{11}(\zeta)]h}{4}.$$

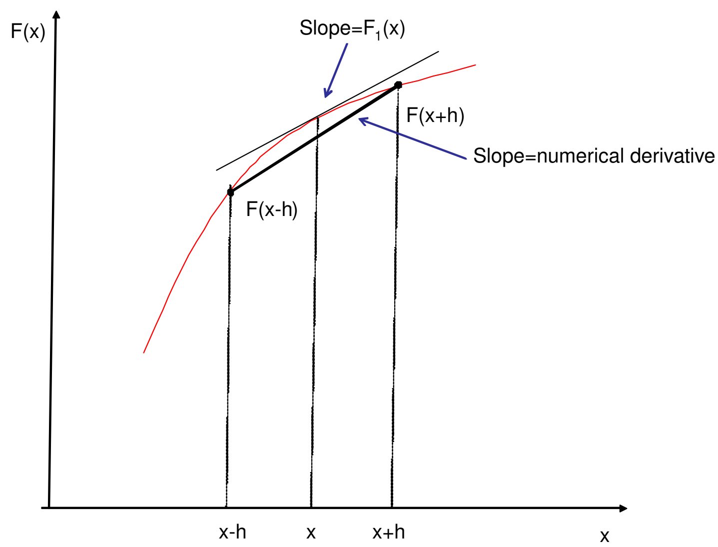


Figure 8.2.1: Finite-difference approximation, $[F(x+h) - F(x-h)]/2h$, vis à vis the true derivative, $F_1(x)$. Mathematically speaking, the error associated with the finite-difference approximation will shrink with h. This ignores the machine error associated with computing $F(x+h)$ and $F(x-h)$.

Now, on the computer the function F will be computed with error, ε. That is, the computer will compute $F(x+h)$ as $F(x+h) + \varepsilon_{+h}$ and $F(x-h)$ as $F(x-h) + \varepsilon_{-h}$. Therefore,

$$F_1(x) = \frac{F(x+h) - F(x-h)}{2h} + \underbrace{\frac{[F_{11}(\xi) - F_{11}(\zeta)]h}{4}}_{\text{approx error}} + \underbrace{(\varepsilon_{+h} - \varepsilon_{-h})/(2h)}_{\text{machine error}}.$$

Thus, there are two types of error on the righthand side of the equation. The mathematical approximation error given by $[F_{11}(\xi) - F_{11}(\zeta)]h/4$ and the machine error shown by $(\varepsilon_{+h} - \varepsilon_{-h})/(2h)$. The first gets smaller when h is reduced while the latter becomes bigger. Note that by computing the derivative in both the forward and backward directions (or centering the derivative around the point x) the mathematical approximation error is reduced.

The second derivative is just the difference between two first derivatives

$$\begin{aligned} F_{11}(x) &= \frac{[F(x+h) - F(x)]/h - [F(x) - F(x-h)]/h}{h} \\ &= \frac{F(x+h) - 2F(x) + F(x-h)}{h^2}. \end{aligned}$$

This second derivative is automatically centered around the point x.

Last, suppose that the function is $F(x, y)$. From the above it is easy to deduce that the numerical derivatives of $F(x, y)$ with respect to y, evaluated at the two points $(x+h, y)$ and $(x-h, y)$, are

$$F_2(x+h, y) = \frac{F(x+h, y+h) - F(x+h, y-h)}{2h}$$

and

$$F_2(x-h, y) = \frac{F(x-h, y+h) - F(x-h, y-h)}{2h}.$$

So, the centered cross derivative is[1]

$$
\begin{aligned}
F_{12}(x,y) &= \frac{[F(x+h,y+h)-F(x+h,y-h)]/2h}{2h} \\
&- \frac{[F(x-h,y+h)-F(x-h,y-h)]/2h}{2h} \\
&= \frac{F(x+h,y+h)-F(x+h,y-h)-F(x-h,y+h)+F(x-h,y-h)}{4h^2}.
\end{aligned}
$$

[1] It should be immediate that

$F_{11}(x,y) = [F(x+h,y) - 2F(x,y) + F(x-h,y)]/h^2$

and

$F_{22}(x,y) = [F(x,y+h) - 2F(x,y) + F(x,y-h)]/h^2$.

8.2.2 *Complex Step Differentiation*

This is a more modern method. It is an accurate method for simple functions–say one line formulas. Here the function F is expanded around the complex part of the point x. Specifically,

$$F(x+ih) \simeq F(x) + F_1(x)ih,$$

where $i \equiv \sqrt{-1}$. Take the imaginary part of both sides and divide by h to get

$$F_1(x) = \frac{\mathrm{Im}(F(x+ih))}{h}.$$

This is a remarkably simple formula and is easy to implement. It turns out to be more accurate than the standard method. Now, h can be smaller, say 1.0e-8.

8.3 *Symbolic Differentiation*

These days computers can manipulate expressions symbolically and come up with analytical solutions, at least when they exist. As an illustration, return to example 2.2 in Chapter 2. The computer will now do the work using MATLAB's symbolic toolbox. The example includes the mathematical expressions written as in Chapter 2, as well as their MATLAB code analogues.

```
% Solving a simple consumption/labor problem analytically using symbolic
% differentiation

clc % clear screen
clear                                   % clear memory
syms h theta w a                        % The symbolic variables

% a = assets
% h = hours worked
% theta is weight on consumption
% w is the wage rate

% Specify the utility function to be maximized
```

The utility function is $\theta \ln(\text{wh} + a) + (1 - \theta)\ln(1 - h)$.

```
utility = theta*log(w*h+a) + (1-theta)*log(1-h);
```

The maximization problem is $\max_h \{\text{theta} * \log(\text{wh} + a) + (1 - \text{theta}) * \log(1 - h)\}$.

```
% Differentiate wrt to h
```

The first-order condition is $\frac{\theta}{\text{wh} + a} w - \frac{(1 - \theta)}{1 - h} = 0$.

```
sol = diff(utility, h); % derivative with respect to h
foc = sol == 0          % first-order condition
```

foc =

$$\frac{\theta\, w}{a + h\, w} - \frac{\theta - 1}{h - 1} = 0$$

```
% Solve first-order condition for h
```

The solution is $h = \theta - (1 - \theta)\frac{a}{w}$.

```
h = solve(foc, h)       % solution for h
```

h =

$$\frac{a\,\theta - a + \theta\, w}{w}$$

```
% Take the derivative of hours worked with respect to wages
```

The derivative is $\frac{\text{dh}}{\text{dw}} = (1 - \theta)\frac{a}{w^2}$.

```
dhdw = diff(h,w)        % dh/dw
```

dhdw =

$$\frac{\theta}{w} - \frac{a\,\theta - a + \theta\,w}{w^2}$$

```
% Simplify the expresson
dhdw = simplify(dhdw)   % simplify expression
```

dhdw =

$$-\frac{a\,(\theta - 1)}{w^2}$$

8.4 Returning to Linear-Quadratic Optimization Problem

Return now to the linear-quadratic optimization problem posed in Chapter 6. Recall that by taking a 2nd-order Taylor expansion the momentary utility function can be expressed as

$$U\Big(F(k)+(1-\delta)k-k'\Big)\simeq U(\widehat{k},\widehat{k}')=\tau+\alpha\widehat{k}-\lambda\widehat{k}'+\rho\widehat{k}\widehat{k}'-\frac{\psi}{2}\widehat{k}^2-\frac{\phi}{2}\widehat{k}'^2,$$

where $\widehat{k}\equiv k-k^*$ and $\widehat{k}'\equiv k'-k^*$. The objective here is to compute the terms $\tau,\alpha,\lambda,\rho,\varphi$, and ϕ using the standard method for numerical differentiation. To do this, let $F(k)=k^\theta$ and $U(c)=\ln(c)$. Set $\delta=0.10$ and $\theta=0.3$. For the step size used for the numerical differentiation, choose $h=0.00001$. Table 8.4.1 compares the analytical and numerical derivatives for the Taylor series expansion. As can be seen, the numerical derivatives match the analytical ones, as least at the 5th decimal point. Figure 8.4.1 shows the quadratic approximation to the utility function.

ANALYTICAL VS NUMERICAL DERIVATIVES

	Analytical	*Numerical*
α	0.95812	0.95812
λ	0.9198	0.9198
ρ	0.94923	0.94923
ψ	0.84603	0.84603
ϕ	0.88128	0.88128

Table 8.4.1: A comparison between the analytical and numerical derivatives for the Taylor expansion.

Recall that the transitional dynamics for the model can be captured using

$$\widehat{k}'=\eta\widehat{k},$$

where η solves the second-order polynomial

$$-\beta\rho\eta^2+(\phi+\beta\psi)\eta-\rho=0.$$

To use the formula input the values obtained for ϕ,ψ and ρ that derive from the quadratic approximation for the utility function–see Table 8.4.1. Let the discount factor, β, equal 0.96. The formula will have two roots. Picking the root with a value less than one (or the stable root) yields $\eta=0.85$. So, each period the fraction 0.15 of the gap between the current capital stock, k, and its steady state level, k^*, is closed. The associated impulse response function is displayed in Figure 8.4.2.

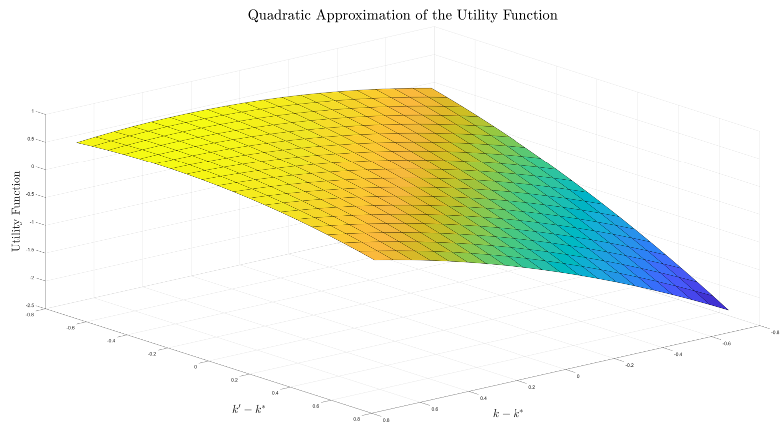


Figure 8.4.1: Quadratic approximation of the utility function. The utility function is increasing in $k - k^*$, decreasing in $k' - k^*$, and convex in both variables.

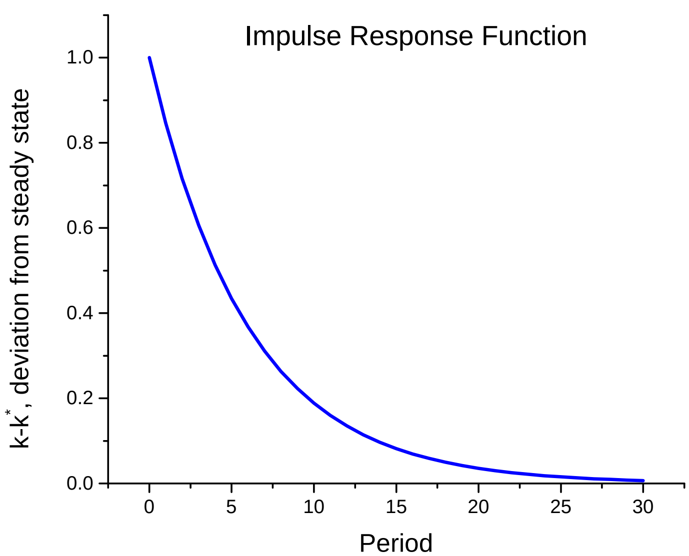


Figure 8.4.2: The capital stock starts off one unit away from the steady state. The impulse response function shows the convergence from this unit deviation to the steady state.

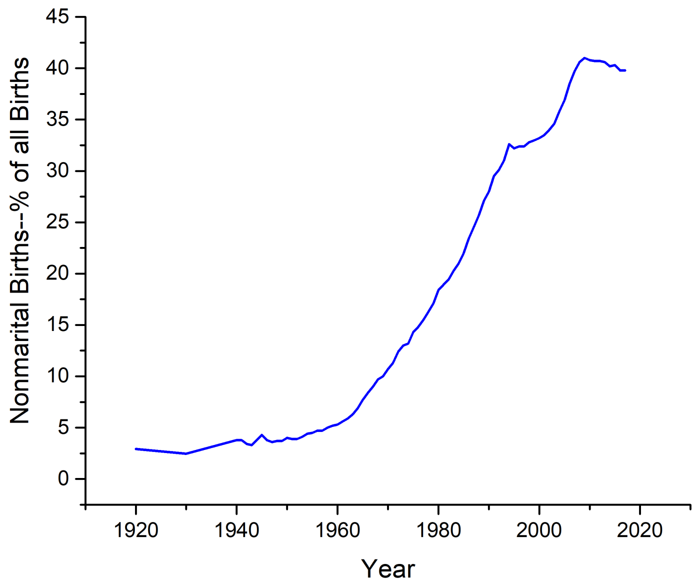


Figure 8.5.1: Non-marital births as a percentage of all births in the United States, 1920-2017. Non-marital births rose despite the fact that contraception became more efficient. *Source*: Greenwood et al. (2021a).

8.5 *The Impact of Technological Innovation in Contraception on Non-Marital Births*

A puzzling feature of the U.S. data is that as contraception improved the number of non-marital births increased–see Figure 8.5.1. Can this be explained? To do this, return to the model of premarital sexual activity presented in Section 3.7 of Chapter 3. Recall that the fraction of 20-year-old women with premarital sexual experience, p, is given by

$$p = \exp[-(\phi O/\eta)^{\beta}],$$

where ϕ is the failure rate of contraception, O is the cost of a non-marital birth, and β and η are the Weibull distribution's shape and scale parameters. When this relationship is calibrated to the U.S. data it transpires that $\beta = 2.30$, $\eta = 2.06$, and $O = 1.34$; again, recall Section 3.7 of Chapter 3. The fraction of women with a non-marital birth, b, is given by

$$b = \phi p = \phi \exp[-(\phi O/\eta)^{\beta}].$$

To calculate non-martial births just multiply the number of sexually active single young women, p, by odds of becoming pregnant, ϕ.

To uncover the impact of the efficacy of contraception on non-marital births, the following derivative and elasticity are computed using complex-step differentiation:

$$\frac{db}{d\phi} \text{ and } \frac{\phi}{b}\frac{db}{d\phi}.$$

Recall the the failure rates for 1900 and 2000 are 72 and 30 percent. The results are shown in the table below. Interestingly, an *drop* in the failure rate for 1900 leads to a *hike* in non-marital births while for 2000 it causes a *fall*. I.e., an one percentage point drop in failure rate in 1900 leads to a 0.47 percentage point *increase* in the fraction of young women having a non-marital birth, while in 2000 it causes a 0.37 percentage point *decrease*. In elasticity terms a one percent drop in ϕ in 1900 leads to 2.18 percent increase in b while for 2000 it causes a 0.48 percent decrease.

CONTRACEPTION AND NON-MARITAL BIRTHS

	$db/d\phi$	$(\phi/b)(db/d\phi)$
1900	-0.47	-2.18
2000	0.37	0.48

What explains this? A drop in the failure rate makes sex safer but it also entices more women to engage in premarital sex. For 1900 the impact on $b = \phi p$ from a rise in p is bigger than the effect from a decline in ϕ. Since not many women were having premarital sex at that time a reduction in the failure rate on non-marital births is modest. For 2000 exactly the opposite is true since a large fraction of unmarried young women were sexually active. As the failure rate approaches 0 so will the number of non-marital births to young women.

8.6 Classical Numerical Intergration

Suppose one wants to compute the function

$$I = \int_{x_0}^{x_n} F(x)dx.$$

That is, the task is to compute the area under the function F on the domain $[x_0, x_n]$, as shown in Figure 8.6.1. The classical way to do this is to break up the distance between x_0 to x_n into a grid of n equally spaced points $\{x_0, x_1, \cdots, x_{j-1}, x_{j,} \cdots, x_n\}$ where $x_j - x_{j-1} = h$ for all $j = 1, \cdots, n$. Now, take any two adjacent points, say x_{j-1} and x_j. Over the interval $[x_{j-1}, x_j]$ the function $F(x)$ will be approximated by a trapezoid, $T^{j-1}(x)$. Specifically,

$$T^{j-1}(x) = (1-\mu)y_{j-1} + \mu y_j, \text{ for } x \in [x_{j-1}, x_j],$$

where

$$y_{j-1} \equiv F(x_{j-1}) \text{ and } y_j \equiv F(x_j),$$

and

$$\mu = (x - x_{j-1})/(x_j - x_{j-1}).$$

Observe that the function $T^{j-1}(x)$ is simply a weighted average of the points $y_{j-1} \equiv F(x_{j-1})$ and $y_j \equiv F(x_j)$ where the weight depends on how close the point is to x_{j-1}. The further x is away from the point x_{j-1}

the higher is the weight that is attached to the point $y_j \equiv F(x_j)$ and consequently the lower is the weight assigned to $y_{j-1} \equiv F(x_{j-1})$. (This is related to the concept of piecewise linear interpolation discussed in Section 8.16.)

Now, the integral for the area under the trapezoid $T^{j-1}(x)$ is

$$\int_{x_{j-1}}^{x_j} T^{j-1}(x)dx = y_{j-1}\int_{x_{j-1}}^{x_j}(1-\frac{x-x_{j-1}}{x_j-x_{j-1}})dx + y_j\int_{x_{j-1}}^{x_j}\frac{x-x_{j-1}}{x_j-x_{j-1}}dx$$
$$= y_{j-1}\int_{x_{j-1}}^{x_j}\frac{x_j-x}{x_j-x_{j-1}}dx + y_j\int_{x_{j-1}}^{x_j}\frac{x-x_{j-1}}{x_j-x_{j-1}}dx$$
$$= \frac{1}{2}y_{j-1}h + \frac{1}{2}y_j h = \frac{y_{j-1}+y_j}{2}h.$$

So, the area of the trapezoid on the interval $[x_{j-1}, x_j]$ is just average of the two points y_{j-1} and y_j multiplied by the length of the interval or h. Summing over all of the trapezoids on the entire domain $[x_0, x_n]$ gives

$$\int_{x_0}^{x_n} F(x)dx \simeq h\sum_{j=1}^{n}\frac{y_{j-1}+y_j}{2}$$
$$= h(y_0/2 + \sum_{j=1}^{n-1} y_j + y_n/2).$$

The accuracy of this approximation will depend on how fine the grid, $\{x_0, x_1, \cdots, x_{j-1}, x_j, \cdots, x_n\}$, is. The more points, or the smaller h is, the higher will be the approximation. Of course, the function $F(x)$

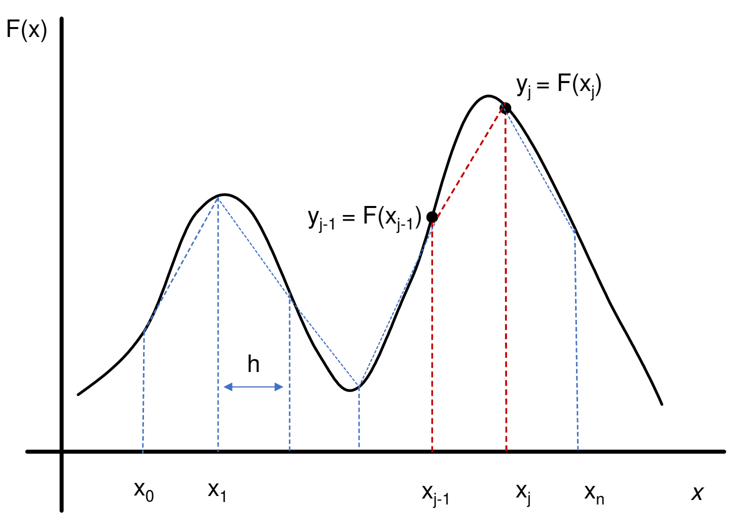


Figure 8.6.1: The area under the curve $F(x)$ between the points x_0 and x_n is approximated by summing the area for a series of n trapezoids.

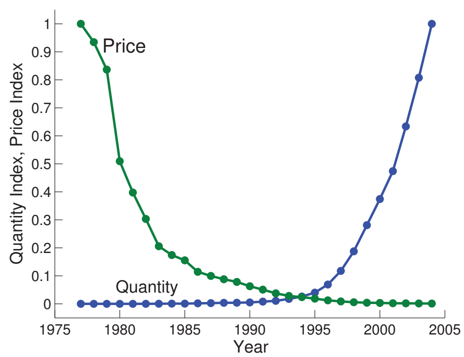


Figure 8.7.1: The plot illustrates the meteoric drop in the quality-adjusted price index for computers and the rocket-like rise in associated quantity index of computers purchased. *Source*: Greenwood and Kopecky (2013)

8.7 *Measuring the Welfare Gain from Personal Computers*

The first PC to be successfully mass produced was the Apple II, which was introduced in 1977. The computer sold for roughly $1,200. Its microprocessor ran at 1MHz and the PC had 4 kb of random-access memory (RAM). There was no hard disk. An audio cassette was used for program loading and data storage. Now, the speed of PCs is measured in terms of gigaHZ, RAM in gigabytes, and hard drive storage in terabytes. Peripherals such as monitors and speakers are also much better. The quality adjusted price of computers dropped at 25 percent per year between 1977 to 2004. Over that time period, spending on computers and peripherals rose from 0 to 0.6 percent of personal consumption spending. The quality-adjusted price and quantity of computers over this period is displayed by Figure8.7.1.

Greenwood and Kopecky (2013) estimated a nonlinear demand curve for computers of the following form for the period 1977 to 2004:

$$c = D(p, y) = \frac{(y + pv)}{p + \theta p^{\rho}} - v, \text{ with } v, \theta, \rho > 0,$$

where c denotes the demand for computers at price p and income y. They find $v = 4.1491 \times 10^{-5}$, $\theta = 0.0056$, and $\rho = 1.4844$. For a given level of income, y, the demand curve for computers has the form shown in Figure 8.7.2. The price p_h where the demand curve hits the vertical axis is known as Hick's virtual price. It solves the equation,

$$\frac{(y + p_h v)}{p_h + \theta p_h^{\rho}} - v = 0.$$

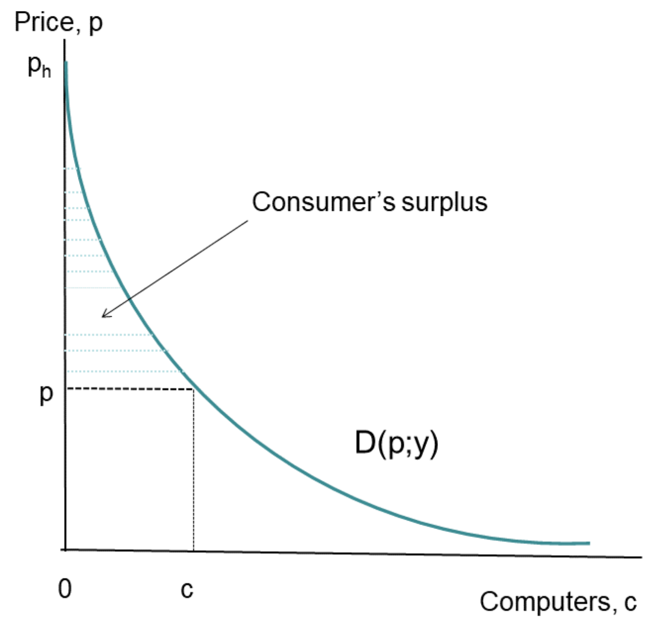


Figure 8.7.2: The demand curve for computers, $c = D(p, y)$. The curve hits the vertical axis at Hick's virtual price, p_h. The area under the demand curve above the price p measures the consumer's surplus. This is the area that needs to be numerically integrated.

The consumer surplus from computers at price p and income y can be estimated by computing the area under the demand curve above the price p. That is,

$$\text{CONSUMER SURPLUS} = \int_{p}^{p_h} [\frac{(y + \widetilde{p}v)}{\widetilde{p} + \theta\widetilde{p}^{\rho}} - v] d\widetilde{p}.$$

This is an exercise in numerical integration. The consumer's surplus at the 2004 indices for price ($p = 0.0013$) and income ($y = 0.2165$) amounts to 2.17 percent of personal consumption expenditure. While consumer's surplus measured in this way is not an exact measure of the worth of computers to consumers, the estimate obtained here is surprisingly close to the compensating and equivalent variations of 2.14 and 2.19 percent reported in Greenwood and Kopecky (2013).

8.7.1 *Welfare Gain from Computers–pseudo code*

1. Define θ, v, ρ, and y_{2004} as global variables.

2. Input in the parameter values for θ, v, ρ,p_{2004}, and y_{2004}.

3. Write a function $C(p)$ for demand for computer that takes the price p as an input and delivers c as output. Pass θ, v, ρ, and y_{2004}into this function as global variables

4. Use a nonlinear equation solver to calculate Hicks virtual price; i.e., find the p_h that solves $C(p_h) = 0$.

5. Calculate consumer surplus by numerically integrating the function $C(p)$ over the range $p = 0$ to $p = p_h$.

6. Divide the answer by y_{2004} and multiply by 100 to express things as a percentage of income.

8.8 *Random Number Generators*

There is nothing random about generating random numbers on a computer. Somewhat surprisingly, these numbers are created using a deterministic algorithm.

8.8.1 *The Modulo Operator*

As a prelude to the discussion, imagine dividing some natural number x through by another natural number $m > 0$ and calculating the remainder, r. (The natural numbers are $0, 1, 2, \cdots$.) The remainder, also a natural number, is given by the formula

$$r = x - \text{FLOOR}(\frac{x}{m})m,$$

where FLOOR is the nearest natural number less than or equal to x/m. The remainder, r, is a natural number. This operation defining the remainder from a division is abbreviated as x modulo m or as

$$x \mod m.$$

Observe that[2]

$$0 \leq r = x \mod m \leq m - 1.$$

Example 8.1. (Remainder from $22 \div 7$) $22 \mod 7 = 22 - \text{FLOOR}(22/7) \times 7 = 22 - \text{FLOOR}(3.1429) \times 7 = 22 - 3 \times 7 = 1$.

Example 8.2. (Remainder from $11 \div 7$) $11 \mod 7 = 11 - \text{FLOOR}(11/7) \times 7 = 11 - \text{FLOOR}(1.5714) \times 7 = 11 - 1 \times 7 = 4$.

Example 8.3. (Remainder from $44 \div 7$) $44 \mod 7 = 44 - \text{FLOOR}(44/7) \times 7 = 44 - \text{FLOOR}(6.2857) \times 7 = 44 - 6 \times 7 = 2$.

[2] To understand the upper bound, observe that for any x the largest value for the remainder must occur when $\text{FLOOR}(x/m) = 0$, which gives $r = x$. But, this requires $m > x$. The largest value of x compatible with this inequality, $m > x$, is $x = m - 1$. The lower bound is easy to deduce. Clearly, r cannot be negative because $x \geq \text{FLOOR}(x/m)m$. But, $\text{FLOOR}(x/m)m = x$ whenever $m = x$. So $0 = x - \text{FLOOR}(x/m)m$ is an admissible value for the remainder when $m = x$.

8.8.2 *The Linear Congruential Generator*

Pick some natural number, m. This is called the modulus of the random number generator. Let x be a natural number that is created using the following iterative scheme

$$x^{j+1} = (ax^j + b) \mod m, \tag{8.8.1}$$

where the constant a, a natural number, is called the multiplier and b, another natural number, is known as the offset coefficient. Given some starting value for x, x^0, this equation will give a sequence of natural numbers. The starting value x^0 is called the seed of the random generator. The natural number x^{j+1} lies in the interval

$$0 \leq x^{j+1} \leq m - 1.$$

If x^{j+1} ever returns a number that was generated before, then the algorithm will repeat itself. Since there are only m possible values of x the generator must cycle after m periods. Define the pseudo-random number, u^{j+1}, that is associated with this sequence by

$$u^{j+1} = \frac{x^{j+1}}{m}.$$

The pseudo-random numbers will resemble those drawn from a uniform distribution–the uniform distribution is discussed in the Chapter A.

Example 8.4. (A 4 period cycle) Let $a = 11$, $b = 0$ and $m = 7$. Set $x^0 = 2$. Then, the algorithm proceeds as follows: $x^1 = (11 \times 2) \mod 7 = 22 \mod 7 = 1$ so that $u^1 = 1/7 = 0.1429$; $x^2 = (11 \times 1) \mod 7 = 4$ so that $u^2 = 4/7 = 0.5714$; $x^3 = (11 \times 4) \mod 7 = 2$ so that $u^3 = 2/7 = 0.2857, \cdots$. Since $x^3 = x^0 = 2$ the random number generator then repeats the cycle over again.

Example 8.5. (A 31 period cycle) Let $a = 13$, $b = 0$ and $m = 31$. Set $x^0 = 3$. Then, the algorithm proceeds as follows: $x^1 = (13 \times 3) \mod 31 = 39 \mod 31 = 8$ so that $u^1 = 8/31 = 0.2581$; $x^2 = (13 \times 8) \mod 31 = 11$ so that $u^2 = 11/31 = 0.3548$; $\cdots$; $x^{30} = (13 \times 5) \mod 31 = 3$ so that $u^{30} = 3/31 = 0.0968, \cdots$. Since $x^{30} = x^0 = 3$ the cycle begins over again.

As can be seen from the above example, for small values of a and m such a random number generator has very poor qualities. But, it works reasonably well for large values. An earlier version of MATLAB set $m = 2^{31} - 1$ (a Mersenne prime number) and $a = 7^5 = 16,807$. This takes $2^{31} - 2$ periods before it cycles.

Eugen Slutsky (1880-1948) was a Russian economist and statistician. He wrote two path-breaking papers in economics. In the the first paper he derived what is now known as the Slutsky equation. This is one of the cornerstones of consumer theory. Because the paper was published in Italian in 1915 it lay in obscurity for a while. Hicks independently rediscovered the notion in 1939. His second paper, the one discussed here, was on "The Summation of Random Causes as a Source of Cyclical Processes." This paper was originally published in Russian in 1927. The English version was published ten years later. The notion of randomness is now everywhere in economics.

8.9 *Eugen Slutsky and Random Causes as the Source of Cyclic Processes*

(I)s it possible that a definite structure of a connection between random fluctuations could form them into a system of more or less regular waves? Many laws of physics and biology are based on chance, among them such laws as the second law of thermodynamics and Mendel's laws. But heretofore we have known how regularities could be derived from a chaos of disconnected elements because of the very disconnectedness. In our case we wish to consider the rise of regularity from series of chaotically-random elements because of certain connections imposed upon them. Eugen Slutsky (1937, p. 106)

In 1937 Eugen Slutsky put forward the following probabilistic model of the business cycle:

$$\begin{aligned} o_t &= x_t + \quad x_{t-1} + \quad \cdots \quad x_{t-9} + \qquad\qquad 5, \\ o_{t-1} &= \qquad\quad x_{t-1} + \quad \cdots \quad x_{t-9} + \quad x_{t-10} + \quad 5, \end{aligned}$$

where o_t is a index of the business cycle for period t and the x_t's are independently and identically distributed random variables. To generate the x_t's, Slutsky took a sample of winning numbers for a lottery for loans from the People's Commissariate of Finance. He just used the last digit of winning numbers for the x_t's. He noted that each of two adjacent values for the business cycle index o, say o_t and o_{t-1}, would have one random cause unique to itself, here x_t and x_{t-10}, and 9 random causes in common, or x_{t-1} to x_{t-9}. Because the business cycle index has events in common there appears to be a correlation between them even though the series of causes are independent. Figure 8.9.1 shows the upshot of Slutsky's Monte Carlo simulation. It exhibits the same pattern of fluctuations as displayed in the business cycle data for ninetieth century England. Think about how much effort and time it

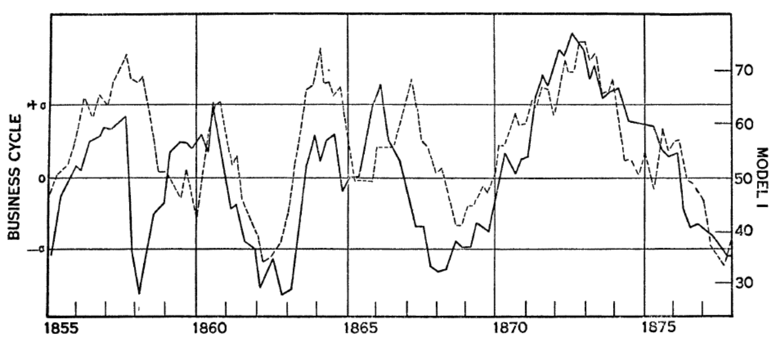


Figure 8.9.1: The dashed line shows an English business cycle index for the years 1855 to 1877. The solid line is a portion of the sample path from Slutsky's Model 1, which corresponds to model shown here. Slutsky cut out a portion of his randomly generated series and aligned it with the business cycle data. *Source*: Slutsky (1937, Figure 3).

would have taken Slutsky do this diagram. This is a simple task today: it involves using a random number generator, a loop for the o_t's, and making a graph from the output.

8.9.1 *Slutsky–pseudo code*

1. Set the seed for random number generator.
2. Call up a sample of uniformly distributed random integers, $\{x_t\}$, where $1 \leq x_t \leq 10$.
3. Iterate on the equation $o_t = x_t + x_{t-1} + \cdots x_{t-9} + 5$, using a for loop.
4. Plot the resulting business cycle. The figure below shows the results.

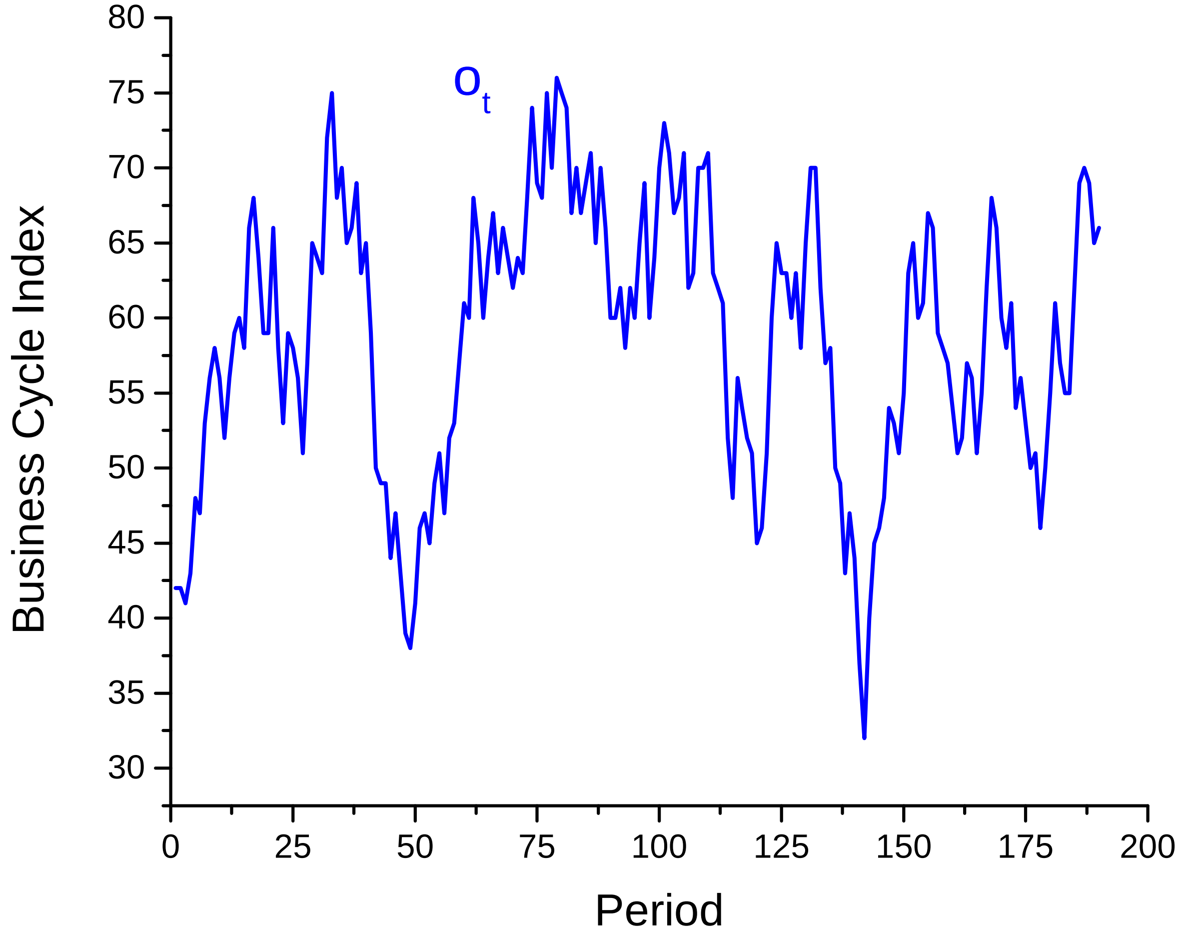


Figure 8.9.2: Slutsky's business cycle à la MATLAB, following the steps in the pseudo code.

8.10 *Monte Carlo Integration*

Again, suppose one wants to compute the function

$$I = \int_a^b F(x)dx.$$

An easy way to do this is by Monte Carlo integration. Now, rewrite the above formula as

$$I = (b-a)[\int_a^b F(x)\frac{1}{b-a}dx].$$

Think about $1/(b-a)$ as representing the density function for a uniform distribution. Thus, the term in brackets is the expected value of the function $F(x)$, while the $(b-a)$ term multiplying this expected value is the length of the line segment going from a to b.

The idea underlying Monte Carlo integration is to compute the expectation term $\int_a^b F(x)/(b-a)dx$ by drawing a random sample of x's on $[a,b]$ from the uniform distribution. Represent this draw of n random numbers by $\{x_1, x_2, \cdots, x_n\}$. The expectation in question is computed using the following formula

$$\int_a^b F(x)/(b-a)dx \simeq \frac{1}{n}\sum_{i=1}^{n} F(x_i).$$

This implies that

$$I = \int_a^b F(x)dx \simeq \frac{(b-a)}{n}\sum_{i=1}^{n} F(x_i),$$

which is the formula used in Monte Carlo integration. Figure 8.10.1 illustrates the situation.

Now, the strong law of large numbers implies that

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n} F(x_i) = \int_a^b F(x)/(b-a)dx,$$

where the righthand side is the expected value of $F(x)$–a statement of the strong law of large numbers is provided in Chapter A. The variance of the sample mean for $F(x)$, or the variance of $\sum_{i=1}^{n} F(x_i)/n$, is given by

$$\sigma_n^2 = \frac{1}{n^2}E[\sum_{i=1}^{n}\{F(x_i) - \sum_{j=1}^{n} F(x_j)\}^2] = \frac{1}{n}E[\{F(x) - E[F(x)]\}^2] = \frac{\sigma^2}{n}.$$

since all the x's are independently and identically distributed. This implies that the sample's standard deviation around the mean, σ_n, will decline with $\sqrt{n}$. Hence, to reduce the standard error by half, the sample size needs to be quadrupled.

8.11 *Robert E. Lucas, Jr., and the Cost of Business Cycles*

What is the welfare cost of random fluctuations in the business activity? This question was poised by Lucas (1987). Surprisingly, in a representative agent framework it is very small. To see this, denote the long-run (or stationary) distribution over consumption, c, and hours worked, h, by $D(c,h)$. In the long-run the distribution over consumption and hours worked is the same in each period.[3] Therefore, expected momentary utility in each period is the same. The representa-

[3] The notion of a long-run distribution is presented later in Section 8.12 on Markov chains. It is also discussed in the Chapter 9.

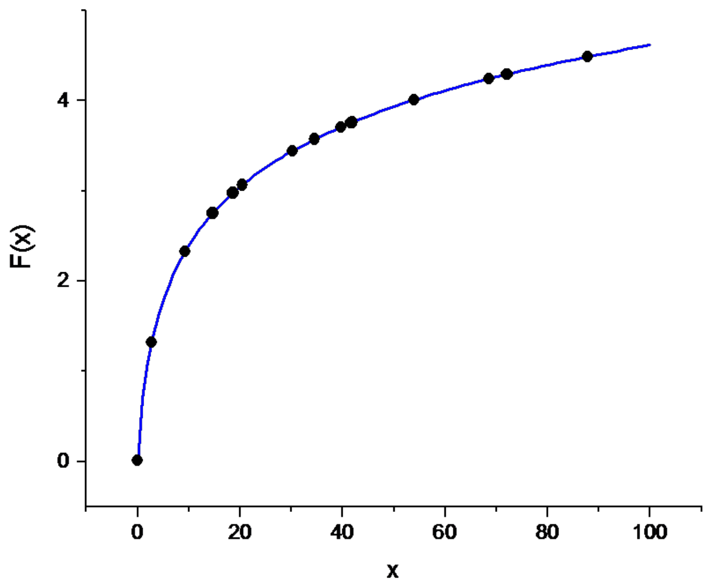


Figure 8.10.1: Monte Carlo Integration. The solid line plots the function $F(x)$. The circles show the $(x_n, F(x_n))$ combinations that arise when the x's are sampled from a uniform distribution and $F(x)$ is evaluated at each of the n sample points. The $F(x_n)$ points will be averaged to compute $E[F(x)]$. Clearly, the larger the sample is the more accurate will be the integration.

tive agent's expected lifetime utility is

$$\begin{aligned}(1+\beta+\beta^2+\cdots)E[U(c)+V(1-h)] &= \frac{E[U(c)+V(1-h)]}{1-\beta} \\ &= \frac{\int\int[U(c)+V(1-h)]dD(c,h)}{1-\beta},\end{aligned}$$

where $U(c)$ is the utility function over consumption, $V(1-h)$ is the one over leisure, and $0<\beta<1$ is the discount factor.

So, what is the welfare cost of business cycles? This involves computing either a compensating or equivalent variation. These concepts are discussed in Chapter 2. Imagine that one could stabilize consumption and hours worked at their respective means, $\bar{c}$ and $\bar{h}$. How much compensation would the representative agent have to be given to make him as well off in the world with business cycle fluctuations as in the world without them? The equivalent variation, ϵ, solves the equation

$$\int\int[U(c(1+\epsilon))+V(1-h)]dD(c,h) = U(\bar{c})+V(1-\bar{h}).$$

The righthand side of the above expression gives the lifetime utility that the representative agent realizes when consumption and hours are stabilized at their mean levels. So, the question being posed is what fraction of consumption in each and every state would the person have to be given to make him as well off as in a world without fluctuations. Observe that there is no β in the formula because it will cancel out of both sides of the equation. Last, despite being complicated looking, this is only one equation in one unknown, ϵ.

To come up with an estimate of the welfare cost of business cycles, let utility be given by

$$U(c) = \theta \ln c \text{ and } V(1-h) = (1-\theta)\ln(1-h),$$

where $\theta = 0.33$. Given the logarithmic form of the utility function, it is easy to see that

$$\theta \ln(1+\epsilon) = \theta \ln(\bar{c}) + (1-\theta)\ln(1-\bar{h}) - \int\int [\theta \ln(c) + (1-\theta)\ln(1-h)]dD(c,h),$$

so that

$$\epsilon = \exp\Big(\{\theta \ln(\bar{c}) + (1-\theta)\ln(1-\bar{h}) - \int\int [\theta \ln(c) + (1-\theta)\ln(1-h)]dD(c,h)\}/\theta\Big) - 1.$$

To use this formula, a distribution over c and h is needed and then a double integration will have to be performed. This will be done using Monte Carlo integration. To obtain the properties of $D(c,h)$ in the U.S. data, time series for consumption, c, and hours, h, are logged and then Hodrick-Prescott (H-P) filtered. The H-P filter is discussed later in this chapter. The standard deviations for consumption and hours are 0.0217 and 0.0257–see Chapter A for a discussion of descriptive statistics. The correlation between these two variables is 0.6160. Figure 8.11.1 shows the bivariate distribution obtained from the U.S. data, for these two variables.

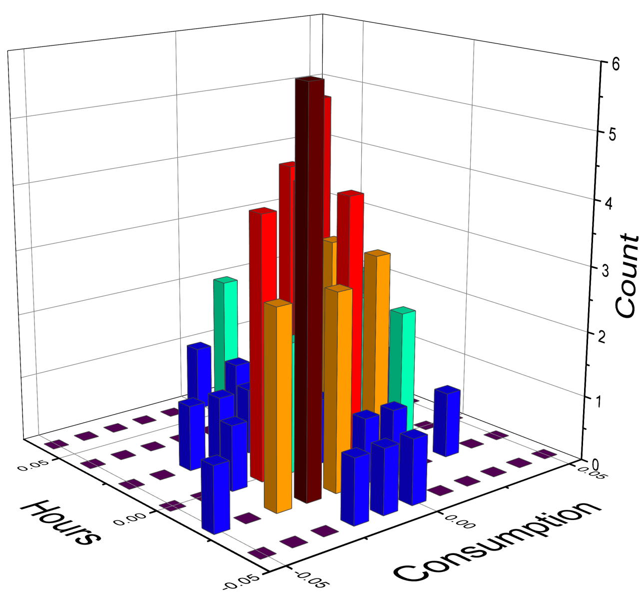


Figure 8.11.1: The bivariate distribution for consumption and hours obtained from annual U.S. data, 1949 to 2017. The data was first logged and then detrended using the Hodrick-Prescott filter. When consumption is high there is a tendancy for hours to also be high, and vice versa–that is, the mass from the bars accumulate along a diagonal in the (c,h) plane from front to back.

For the analysis assume that $\ln c$ and $\ln h$ follow the bivariate normal distribution $N(\mu_{\ln c}, \mu_{\ln h}, \sigma^2_{\ln c}, \sigma^2_{\ln h}, \sigma_{\ln c, \ln h})$–the bivariate normal distribution is presented in Chapter A. As in the U.S. data, set the variances and covariance as follows: $\sigma^2_{\ln c} = 0.0217^2, \sigma^2_{\ln h} = 0.0257^2$,and $\sigma_{\ln c, \ln h} = 0.6160 \times (0.0217 \times 0.0257)$. Without loss of generality, pick $\mu_{\ln c} = -0.0002$ and $\mu_{\ln h} = -1.0989$, which imply that the means of c and h are 1.0 and 0.3333. Call up 100,000 random draws of (c, h) from this distribution and then calculate

$$\frac{1}{100,000} \sum_{i=1}^{100,000} [\theta \ln(c_i) + (1-\theta) \ln(1-h_i)],$$

which is a Monte Carlo approximation to the expectation $\int \int [\theta \ln c + (1-\theta) \ln(1-h)] dD(c, h)$. Using the expectation in the formula for ϵ results in $100 \times \epsilon = 0.040$. Thus, the representative agent needs be given an extra 0.04 percent in terms of consumption to make him as well off as in a world without business cycles! By comparison, Lucas (1987, Table 2, p. 26)'s estimate of the cost of consumption fluctuations was 0.072 percent, at least when the utility function was logarithmic and consumption had a standard deviation of 0.039. Lucas ignored variability in labor supply. Hours worked tend to be low when consumption is low. Hence, leisure is negatively associated with consumption which partially offsets the loss from a fall in consumption. So, adding variability in labor supply works to reduce the cost of business cycles. His conclusion was that the welfare cost of business cycles is very low, especially in comparison with the large welfare effects associated with changes in an economy's growth rate–recall the discussion in Chapter 2.

8.11.1 *Lucas–pseudo code*

1. Load data on consumption and hours worked.

2. Log and HP filter the data–the Hodrick-Prescott filter is discussed in Section 8.17.

3. Calculate business cycle statistics from detrended logged data: variances of consumption and hours and the covariance between the two series.

4. Set the seed for the random number generator.

5. Call up a sample of multivariate normally distributed random variables, $\{\ln c_t, \ln h_t\}$, plugging in the variance-covariance matrix estimated from the data.

6. Unlog consumption and hours worked.

7. Compute $W^B = \theta \ln(\sum_{i=1}^{100,000} c_t/100,000) + (1-\theta) \ln(1 - \sum_{i=1}^{100,000} h_t/100,000)) = \theta \ln(\bar{c}) + (1-\theta) \ln(1-\bar{h})$.

8. Compute expected utility as given by $W^A = \sum_{i=1}^{100,000} [\theta \ln(c_i) + (1 - \theta)\ln(1 - h_i)]/100,000$.

9. Calculate the equivalent variation in percentage terms using the formula $100 \times \{\exp[(W^B - W^A)/\theta] - 1\}$.

8.12 *Markov Chains*

A Markov chain is a probability model where the system jumps from one state to another in a random manner. The odds of how the next jump will occur depend only on the current state of the system. Suppose the system can take one of n values at each point in time, denoted by $z \in \mathcal{Z} \equiv \{z_1, z_2, \cdots, z_n\}$, where the set $\mathcal{Z}$ is time invariant. If the system is currently at state z_i, then the chance of hopping next period to state z_j is given by

Andrey A. Markov (1856 -1922) was a Russian mathematician. He is best known for his work on stochastic processes.

$$\pi_{ij} = \Pr[z' = z_j | z = z_i], \text{ for all } i, j = 1, \cdots, n.$$

The π_{ij}'s are called *transition probabilities*. Since the odds of how the system moves depend solely on where the system is currently, the system is *memoryless*. Now,

$$\sum_{j=1}^{n} \pi_{ij} = 1, \text{ for all } i,$$

because if the system is currently at z_i, it must either stay next period at z_i or move to some z_j for $j = 1, \cdots, n$ with $j \neq i$.

Example 8.6. (Two-State Markov Chain) Let z have two values, a low value (state 1) and a high one (state 2) represented by $\underline{z}$ and $\overline{z}$. Suppose that the odds of going from the low value (state 1) to the high value (state 2) are given by π_{12} and from the high value to the low value are π_{21}. These are called transition probabilities. Note that $\pi_{11} = 1 - \pi_{12}$ and $\pi_{22} = 1 - \pi_{21}$. Figure 8.12.1 illustrates the situation. A common assumption is to set $\pi_{12} = \pi_{21}$. This implies $\pi_{11} = \pi_{22}$. This renders the Markov chain is symmetric.

Now, load the transition probabilities, or the π_{ij}'s, into what as known as a *transition matrix*, T, as follows

$$\underbrace{T}_{n \times n} = \begin{bmatrix} \pi_{11} & \cdots & \pi_{1n} \\ \vdots & \ddots & \vdots \\ \pi_{n1} & \cdots & \pi_{nn} \end{bmatrix}.$$

Each row of the matrix sums to one, since $\sum_{j=1}^{n} \pi_{ij} = 1$, for all i. Suppose that one is given some initial probability distribution, ρ^0, over the position of the states, where

$$\underbrace{\rho^0}_{1 \times n} = (\rho_1^0, \rho_2^0, \cdots, \rho_n^0).$$

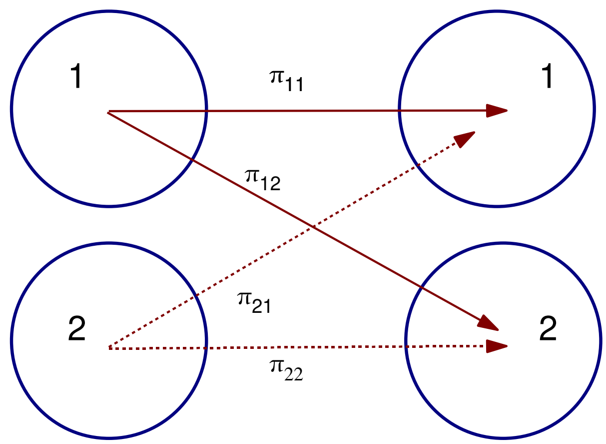


Figure 8.12.1: Two-State Markov Chain

Next period's probability distribution over the states, or ρ^1, is given by

$$\underbrace{\rho^1}_{1\times n} = \underbrace{\rho^0}_{1\times n} \times \underbrace{T}_{n\times n}.$$

If one knew with certainty that the time-0 state of the system is z_i, then just set $\rho_i^0 = 1$ and $\rho_j^0 = 0$ for all $j \neq i$. So, writing the time-0 state of the system in the above probabilistic manner is without loss of generality.

By expanding the above formula, it can be seen that

$$(\rho_1^1, ..., \rho_n^1) = (\rho_1^0, ..., \rho_n^0) \begin{bmatrix} \pi_{11} & \cdots & \pi_{1n} \\ \vdots & \ddots & \vdots \\ \pi_{n1} & \cdots & \pi_{nn} \end{bmatrix} = (\sum_{i=1}^{n} \rho_i^0 \pi_{i1}, \cdots, \sum_{i=1}^{n} \rho_i^0 \pi_{in}).$$

Take the probability of being in state j next period or ρ_j^1. This is given by $\rho_j^1 = \sum_{i=1}^{n} \rho_i^0 \pi_{ij}$. In theory one can get to state j in period 1 by starting out from any state i in period 0. The odds of initially being in state i are given by ρ_i^0 while the odds of then transiting from i to j are π_{ij}. So, $\rho_i^0 \pi_{ij}$ represents the probability of starting out in i and then moving to j. Potentially the system can get into state j period in period 1 from any period-0 state i, so sum over all initial positions to get $\rho_j^1 = \sum_{i=1}^{n} \rho_i^0 \pi_{ij}$.

8.12.1 Stationary Distribution

It easy to see that the m-period-ahead probability distribution over states is given by

$$\rho^m = \rho^{m-1}T = \rho^{m-2}T \times T = \rho^{m-2}T^2 = \cdots = \rho^0 T^m.$$

A question of interest is whether or not ρ^m converges to something as m gets large. The long-run or stationary distribution, ρ^*, is given by the fixed point to this operation or

$$\rho^* = \rho^* T. \tag{8.12.1}$$

There are five ways to compute the stationary distribution.

1. The easiest way to compute ρ^* is to iterate on the mapping

$$\rho^{j+1} = \rho^j T, \tag{8.12.2}$$

until $|\rho^{j+1} - \rho^j|$ becomes sufficiently small. At each stage, the vector ρ^j is just post multiplied by the matrix T. This is an easy thing to do.

2. You can also rise the matrix T to a large power. Each row of the corresponding matrix converges to the stationary distribution as $T \to \infty$. On this, it is easy to see that

$$(\rho_1^*, ..., \rho_n^*) = (\rho_1^*, ..., \rho_n^*) \begin{bmatrix} \rho_1^* & \cdots & \rho_n^* \\ \vdots & \ddots & \vdots \\ \rho_1^* & \cdots & \rho_n^* \end{bmatrix}.$$

3. Alternatively, solve the following system of equations for ρ^*:

$$\underbrace{(\rho_1, ..., \rho_{(n-1)}, \rho_n)}_{\rho^*} = (\rho_1, ..., \rho_{(n-1)}, \rho_n) \underbrace{\begin{bmatrix} \pi_{11} & ... & \pi_{1(n-1)} & -1 \\ \vdots & & & \vdots \\ \pi_{(n-1)1} & & \pi_{(n-1)(n-1)} & -1 \\ \pi_{n1} & ... & \pi_{n(n-1)} & 0 \end{bmatrix}}_{\widehat{T}}$$

$$+ \underbrace{(0, ..., 0, 1)}_{b},$$

or

$$\rho^* = b * [I - \widehat{T}]^{-1}. \tag{8.12.3}$$

Since the matrix $I - T$ does not have full rank, the last equation in the system (or the one for ρ_n) is replaced by the equation $\sum_i \rho_i = 1$, which can be written as $\rho_n = 1 - \sum_{i \neq n} \rho_i$. This is resolved in the above system of equations by writing the last equation as $\rho_n = -\rho_1 \cdots - \rho_{n-1} + 1$. (To see that $I - T$ does not have full rank replace each π_{ii} with $1 - \sum_{j \neq i} \pi_{ij}$ $(= \pi_{ii})$. The sum of the last $n-1$ columns in $I - T$ then equals the negative of the first one.)

4. Yet another way is to compute the eigenvalues and (left) eigenvectors associated with the transition matrix T. An eigenvalue/eigenvector pair must solve the equation

$$\underbrace{e}_{1\times n}\underbrace{T}_{n\times n} = \varepsilon \underbrace{e}_{1\times n},$$

where e is an $1 \times n$ eigenvector and ε its associated eigenvalue, which is a scalar–see Chapter A. Now, the stationary distribution solves this equation when $\varepsilon = 1$ and $e = \rho^*$ (see equation (8.12.1) above). So, one just needs to find the eigenvector associated with an eigenvalue of one. This can be achieved by factorizing matrix T as $T = E\Lambda E^{-1}$, where each column of E is an eigenvector e. Through this eigendecomposition of T, the diagonal elements of the diagonal matrix Λ are the eigenvalues ε associated with each eigenvector. Note that if e solves $eT = \varepsilon e$, then so will $(\xi e)T = \varepsilon(\xi e)$, where ξ is a scalar. Choose ξ to normalize the eigenvector e, associated with $\varepsilon = 1$, so that $\sum_{i=1}^{n} e_i = 1$.

5. Last, one could conduct a Monte Carlo simulation to find the stationary distribution. To do this, the Markov chain is simulated for a long time series of identically and independently distributed random shocks drawn from a uniform distribution on $[0, 1]$. Suppose in some time period that the Markov chain is in state i. The Markov chain will move to state j next period if the shock for this period lies in the interval $[\sum_{l=1}^{j} \pi_{il}, \sum_{l=j}^{j+1} \pi_{il}]$.

One might guess that the existence of a unique, invariant long-run distribution might be related to whether or not the operator T is a contraction mapping–see Chapter 6 for the definition of a contraction mapping. Let $\mathcal{P}^n$ represent the space of n-dimensional probability vectors. Think about the transition matrix as defining an operator $T : \mathcal{P}^n \to \mathcal{P}^n$.

Lemma 8.1. *(Convergence to a unique, invariant long-run distribution)* $\lim_{m\to\infty} \rho^0 T^m = \rho^*$ *for all* $\rho^0 \in \mathcal{T}^n$ *if and only if for some* m *it occurs that* T^m *defines a contraction on* $\mathcal{P}^n$.

Proof. See Stokey and Lucas (1986, chp 11). □

A necessary and sufficient condition for this to occur is given next – again see Stokey and Lucas (1986, chp 11). In more conventional fashion, now let

$$T = \underbrace{[\pi_{ij}]}_{n\times n}.$$

Condition 8.1. (Mixing) For some j there exists a m such that $\min_i \pi_{ij}^m > 0$.

That is, there exists a *column* where all the entries have non-zero elements. This condition implies that at iteration m it is possible to get into state j from any other state. In other words, the system can't get stuck with probability one in any other state $i \neq j$.

Some examples for two-state Markov chains are presented now.

Example 8.7. (Long-run distribution for a two-state Markov chain.) Let $T = \begin{bmatrix} \pi_{11} & 1-\pi_{11} \\ 1-\pi_{22} & \pi_{22} \end{bmatrix}$. The stationary distribution, if it exists, must solve

$$(\rho_1, \rho_2) = (\rho_1, \rho_2) \begin{bmatrix} 1-\pi_{12} & \pi_{12} \\ \pi_{21} & 1-\pi_{21} \end{bmatrix}.$$

It's easy to check that

$$\rho_1^* = \pi_{21}/(\pi_{12} + \pi_{21}) \text{ and } \rho_2^* = \pi_{12}/(\pi_{12} + \pi_{21})$$

satisfy this equation. Or one could use the equations $\rho_1^* = \rho_1^*(1 - \pi_{12}) + \rho_2^* \pi_{21}$ and $\rho_1^* + \rho_2^* = 1$, which is really just a restatement of (8.12.3) for the 2×2 case.

Example 8.8. (Long-run variance and autocorrelation for a symmetric two-state Markov chain) Let $z \in \{-\delta, \delta\}$. Define π by $\pi \equiv \pi_{11} = \pi_{22}$. From the previous example it is clear that $\rho_1^* = \rho_2^* = 0.5$. The long-run mean of shock is $E[z] = -\delta/2 + \delta/2 = 0$. It is easy to see that $E[z^2] = \delta^2/2 + \delta^2/2 = \delta^2$. Additionally, $E[z]^2 = 0$. Thus, variance and standard deviation are given by δ^2 and δ. Finally, $E[z'z] = (\pi_{11} - \pi_{12} - \pi_{21} + \pi_{22})\delta^2/2 = (2\pi - 1)\delta^2$. This implies that the long-run coefficient of autocorrelation is $(2\pi - 1)\delta^2/\delta^2 = 2\pi - 1$.

It may be the case that a unique invariant long-run distribution does not exist, as the following two examples show. Hence, they can't satisfy the above mixing condition.

Example 8.9. (A long-run distribution does not exist) Let $T = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$. Check that with this transition matrix, if you start out in state 1, you'll switch to state 2 and vice versa. Here, it is easy to deduce that $\lim_{m\to\infty} \rho^0 T^m$ doesn't exist for certain ρ^0–try $\rho^0 = [1, 0]$. It is easy to see that $T^m = T$ for all odd m and $T^m = I$ for all even m. Hence, the mixing condition will never obtain.

Example 8.10. (The long-run distribution is not invariant) Let $T = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$. Check that with this transition matrix, if you start out in state 1 you'll stay there, while the same is true for state 2. Pick any conjectured ρ^*. The above fact implies that $\lim_{m\to\infty} \rho^0 T^m \neq \rho^*$for *all*

$\rho^0 \in \mathcal{P}^2$. Here there are different ρ^* associated with different starting values, ρ^0–try $\rho^0 = [1, 0]$ and $\rho^0 = [0, 1]$. Finally, observe that $T^m = I$ for all m, so the mixing condition cannot ever be satisfied.

8.12.2 *Monte Carlo Simulation of a Markov Chain*

A Markov chain can be simulated using a Monte Carlo procedure. This is often useful when an economic model involves a Markov chain. To do this, draw a sample of m uniformly distributed random variables, $\{\varepsilon_t\}_{t=1}^m$. Each $\varepsilon_t \in [0, 1]$. Now, iterate along a sequence starting at $t = 1$ and going to $t = m$. At iteration t suppose that $z_{t-1} = z_r \in \mathcal{Z}$. The shock may randomly transit to another value $z_t = z_s$ in this iteration. Now, take ε_t from the sample of random variables. Compute the current technology shock, z_t, follows:

$$z_t = z_s, \text{ if } \varepsilon \in [\sum_{u=0}^{s-1} \pi_{r,u}, \sum_{u=0}^{s} \pi_{r,u}], \text{ for } s = 1, \cdots, n,$$

where $\pi_{r0} \equiv 0$ and $\sum_{u=1}^n \pi_{ru} = 1$. This is portrayed for a 3-state Markov chain in Figure 8.12.2. From this a sequence for the z's can be obtained, $\{z\}_{t=1}^m$. To start the simulation, z_0 can just be set to some $z \in \mathcal{Z}$. The starting value won't matter much when m is large. The obtained sequence for the z's, or $\{z\}_{t=1}^m$, can be used to simulate a model involving a Markov chain.

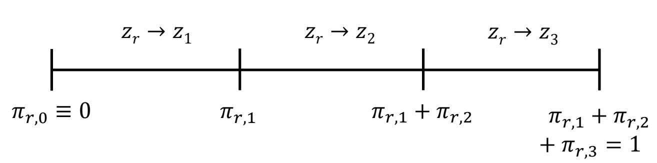


Figure 8.12.2: Monte Carlo simulation of the shocks, z_t, for a 3-State Markov chain. Suppose initially that $z_{t-1} = z_r$. The system can move in period t to either z_1, z_2 or z_3. Split the unit line $[0, 1]$ into three intervals, $u = 1, 2, 3$. The length of interval $u = 1, 2, 3$ is equal to $\pi_{r,u}$. The total length over all intervals is 1. If the value of $\varepsilon_t \in [0, 1]$ lies in interval u, then the shock transits between iterations $t - 1$ and t from z_r to z_u.

8.13 Unemployment: Calibrating a Markov Chain

Many workers in the United States move back and forth between employment and unemployment. The pool of unemployed depends on the flow in from workers who got separated from their jobs and the flow out from the unemployed who found jobs. The goal here is represent this process using a two-state Markov chain over the employed and unemployed states. The Markov chain will be calibrated so that it matches the long-run US data on the average rate of unemployment and the average duration of unemployment. This involves backing out the job finding and separation rates in the transition matrix for the Markov chain so that it matches these long-run facts.

Let the transition between employment, e, and unemployment, u, be governed by the job-finding probability ϕ—how likely an unemployed worker finds a job—and the separation probability σ—how likely an employed worker becomes unemployed. Given this, the aggregate rates of employment and unemployment will evolve according to

$$e_{t+1} = (1-\sigma)e_t + \phi u_t$$

and

$$u_{t+1} = \sigma e_t + (1-\phi)u_t.$$

These equation are easy to explain. Take the one for unemployment. In period t the employment rate is e_t. Out of this component of the labor force the fraction σ will lose their job and enter unemployment in period $t+1$. Likewise, in period t the unemployment rate is u_t. From pool of unemployed the fraction $1-\phi$ will not find a job and therefore will remain unemployed in period $t+1$. An interesting statistics to analyze is the average duration of unemployment, d. This is given by[4]

$$\begin{aligned} d &= 1\phi + 2(1-\phi)\phi + 3(1-\phi)^2\phi + \cdots + n(1-\phi)^{n-1}\phi + \cdots \\ &= \phi\left(1 + 2(1-\phi) + 3(1-\phi)^2 + \cdots + n(1-\phi)^{n-1} + \cdots\right) \\ &= 1/\phi. \end{aligned}$$

[4] To go from the penultimate to the last line, recall the formula for a geometric series, or $1/(1-x) = 1 + x + x^2 + x^3 + \cdots$,where $0 < x < 1$. Differentiate both sides with respect to x to get $1/(1-x)^2 = 1 + 2x + 3x^2 + \cdots$. Now, set $x = 1-\phi$ so that $1-x = \phi$.

To understand this formula, note that to be unemployed for exactly n periods, you must have not found a job for $n-1$ consecutive periods, which occurs with probability $(1-\phi)^{n-1}$, and then found a job at the end of the n^{th} period, which happens with odds ϕ.

The stationary distribution for employment and unemployment must solve the following version of (8.12.3),

$$(e, u) = (e, u)\begin{bmatrix} 1-\sigma & -1 \\ \phi & 0 \end{bmatrix} + (0, 1),$$

from which is easy to calculate that[5]

[5] Again, this amounts to using the two equations $e = e(1-u) + u\phi$ and $e + u = 1$.

$$e = \frac{\phi}{\sigma + \phi} \text{ and } u = \frac{\sigma}{\sigma + \phi}.$$

To calibrate the above Markov chain two facts will be used. First, the long-run rate of unemployment is set to 5.77 percent, the average monthly unemployment rate since 1948, implying that $u = 0.0577$. Second, the average unemployment duration is taken to be 16.2254 weeks so that $d = 16.2254/4 = 4.0564$ months. Both trends are shown in Figure 8.13.1. The job finding probability is then given by $\phi = 1/4.0564 = 0.2465$, so that 24.65 percent of the unemployment find a job every month. From the formula for the long-run unemployment, it follows that the job separation rate is $\sigma = u\phi/(1-u) = 0.0151$. Thus, 1.51 percent of workers are separated monthly from their jobs.

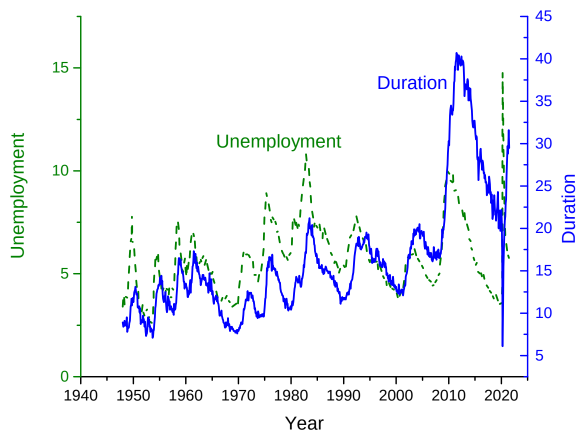


Figure 8.13.1: The unemployment rate (% of labor force) and average unemployment duration (in weeks) over the last 60 years.

8.14 *The Equity Premium: A Puzzle*

From 1889-1978 the average return on equity from the Standard and Poor 500 index was 7 percent. In contrast, the average yield on short term debt was less than 1 percent. Can such a differential be explained within the neoclassical growth model? The puzzle is that to get a low risk-free interest rate in a growing economy you need a high elasticity of intertemporal substitution. When future income is higher than current income individuals would like to borrow. This operates to drive up the interest rate. To mitigate this, people must be very willing to postpone consumption in response to interest rate rises; in other words, a high elasticity of intertemporal substitution is required. To get a large equity premium, people must dislike risk. This requires a high coefficient of relative risk aversion. But in a standard macro model, with isoelastic preferences for consumption, the coefficient of relative risk aversion (discussed below) is the reciprocal of the elasticity of intertemporal substitution (discussed in Chapter 6). In a classic paper, Mehra and Prescott (1985) show that the standard macro model can only generate an equity premium that is 0.4 percentage points higher than on short-term debt.

8.14.1 *The Setup*

Consider a representative agent economy where the individual has tastes of the following form

$$U(c, \alpha) = \frac{c^{1-\alpha} - 1}{1 - \alpha},\ 0 < \alpha < \infty.$$

where c is consumption. Now, $1/\alpha$ is the elasticity of intertemporal substitution–see Chapter 6 for discussion about the elasticity of intertemporal substitution. With this utility function α also represents the coefficient of relative risk aversion, as is discussed below. So, α has a double duty. This is at the heart of the equity premium puzzle. The person discounts the future at rate β. Let the person's income, y, evolve according to the two-state Markov chain

$$y' = zy,$$

where $z \in \{z_1, z_2\}$ with $\pi_{ij} \equiv \Pr[z' = z_j | z = z_i]$. Now, assume that

$$z_1 = 1 + \mu + \delta \text{ and } z_2 = 1 + \mu - \delta.$$

Here μ governs how consumption grows while δ controls its volatility. Additionally, assume that

$$\pi_{11} = \pi_{22} \equiv \pi,$$

which implies that $\pi_{12} = \pi_{21} = 1 - \pi$. Thus, the Markov chain is symmetric.

Definition 8.1. (Coefficient of relative risk version) The coefficient of relative risk aversion is defined as $\theta = -c[U_{11}(c)/U_1(c)]$. With the utility function $U(c, \alpha) = (c^{1-\alpha} - 1)/(1 - \alpha)$ it is clear that $\alpha = \theta$. To understand this concept, consider a static setting where an individual may invest a fraction, ϕ, of his wealth, c, in a risky asset that will payoff either $\gamma + \varepsilon$ or $\gamma - \varepsilon$ with probability $1/2$, where $\gamma > 1$ and $\varepsilon > 0$. So, the expected payoff on a unit of investment in this risky asset is $\gamma > 1$. At that time of the payoff the person will consume all of his wealth. So, the individual's problem is

$$\max_{\phi}\{U((1-\phi)c + \phi c(\gamma + \varepsilon)) + U((1-\phi)c + \phi c(\gamma - \varepsilon))\}/2.$$

The first-order condition associated is

$$U_1((1-\phi)c + \phi c(\gamma+\varepsilon)) \times (\gamma+\varepsilon-1) + U_1((1-\phi)c + \phi c(\gamma-\varepsilon)) \times (\gamma-\varepsilon-1) = 0.$$

Now, take a first-order Taylor expansion of the above two marginal utility terms around c to get

$$U_1(c)\times(\gamma+\varepsilon-1)+U_{11}(c)\phi c(\gamma+\varepsilon-1)^2/2+U_1(c)\times(\gamma-\varepsilon-1)+U_{11}(c)\phi c(\gamma-\varepsilon-1)^2/2=0.$$

Next, let $\varepsilon \to 0$, which amounts to assuming that the risk is small. Then, it transpires that

$$\phi=\frac{1}{\theta}=\frac{1}{\alpha}=-\frac{1}{c}\frac{U_1(c)}{U_{11}(c)}.$$

So, the reciprocal of the coefficient of relative risk aversion can be thought of as measuring the fraction of wealth that an individual will invest in a risky asset. The bigger the coefficient of relative risk aversion, θ, the smaller will be the amount invested in the risky asset.

8.14.2 *Pricing Equities and Bonds*

The prices for equities and bonds will now be characterized. Since this is a representative agent model, there will never be any trades in either equities or bonds. That is, the person will always consume his endowment in a period. Equities and bonds are priced so that there will always be zero stock and bond trades in equilibrium.

Equity

Suppose that an equity is a claim on the flow of income, y. Let the price of a share be denoted by p. This price will be a function of the state of the economy So, let $p = P(y, z_i)$ represent the current price of equity and $p' = P(y', z_j) = P(z_j y, z_j)$ denote next period's price. If the individual buys a share in the current period, his consumption will be reduced by $P(y, z_i)$. The marginal utility of current consumption is $y^{-\alpha}$, so his utility will be reduced by $y^{-\alpha}P(y, z_i)$. Next period the person gets a dividend in the random amount y' and will be able to sell the share at the price $P(z_j y, z_j)$. Thus, utility next period will be increased by the random amount $\beta y'^{-\alpha}[y' + P(y', z_j)] = \beta(z_j y)^{-\alpha}[z_j y + P(z_j y, z_j)]$. This event occurs with chance π_{i1}. So, the person's Euler equation is

$$y^{-\alpha}P(y,z_i)=\pi_{i1}\beta(z_1y)^{-\alpha}[z_1y+P(z_1y,z_1)]+\pi_{i2}\beta(z_2y)^{-\alpha}[z_2y+P(z_2y,z_2)],$$

or

$$P(y,z_i)=\pi_{i1}\beta(z_1)^{-\alpha}[z_1y+P(z_1y,z_1)]+\pi_{i2}\beta(z_2)^{-\alpha}[z_2y+P(z_2y,z_2)].$$

Now, conjecture that the pricing function is given by $P(y, z_i) = w_i y$ implying that $P(z_j y, z_j) = w_j z_j y$. This guess looks reasonable because the above equation is linear in y and any functional dependence on z_i

can be captured by the constant w_i; i.e., think about writing $w_i = Z(z_i)$ where Z is some function. If so, then

$$w_i = \beta\pi_{i1}z_1^{1-\alpha}(1+w_1) + \beta\pi_{i2}z_2^{1-\alpha}(1+w_2), \text{ for } i = 1,2.$$

This is a system of two equation in two unknowns and can be represented in matrix notation by

$$w = \beta\Lambda w + \gamma,$$

where

$$w = \begin{bmatrix} w_1 \\ w_n \end{bmatrix}, \Lambda = \begin{bmatrix} \pi_{11}z_1^{1-\alpha} & \pi_{12}z_2^{1-\alpha} \\ \pi_{21}z_1^{1-\alpha} & \pi_{22}z_2^{1-\alpha} \end{bmatrix}, \gamma = \begin{bmatrix} \beta(\pi_{11}z_1^{1-\alpha} + \pi_{12}z_2^{1-\alpha}) \\ \beta(\pi_{21}z_1^{1-\alpha} + \pi_{22}z_2^{1-\alpha}) \end{bmatrix}.$$

Thus,

$$w = [I - \beta\Lambda]^{-1}\gamma,$$

assuming that $|I - \beta\Lambda| \neq 0$.

What is the *expected* return from holding equity? The *realized* return, r_{ij}, when moving from state (y, z_i) to $(z_j y, z_j)$ is

$$r_{ij} = \frac{P(z_j y, z_j) + z_j y - P(y, z_i)}{P(y, z_i)} = \frac{z_j(w_j + 1)}{w_i} - 1.$$

The expected return on equity, conditional on that the current state is i, is

$$R_i = \pi_{i1}r_{i1} + \pi_{i2}r_{i2}.$$

Thus, the long-run return on equity is

$$R^e = \rho_1^* R_1 + \rho_2^* R_2.$$

Bonds

Next consider the price of a one-period discount bond in state i, or $p_i^f = P^f(y, z_i)$. Such a bond will pay off one unit of consumption next period with certainty. Even so, the marginal utility of next period's consumption is a random variable dependent on the individual's income. The Euler equation for the one-period discount bond reads

$$y^{-\alpha}P^f(y, z_i) = \pi_{i1}\beta(z_1 y)^{-\alpha} + \pi_{i2}\beta(z_2 y)^{-\alpha},$$

so that

$$P^f(y, z_i) = \frac{\pi_{i1}\beta(z_1 y)^{-\alpha} + \pi_{i2}\beta(z_2 y)^{-\alpha}}{y^{-\alpha}} = \beta(\pi_{i1}z_1^{-\alpha} + \pi_{i2}z_2^{-\alpha}).$$

The expected return on this risk free asset, conditional on that the current state is i, is

$$R_i^f = 1/P^f(y, z_i) - 1,$$

which implies that the long-run return will be

$$R^f = \rho_1^* R_1^f + \rho_2^* R_2^f.$$

8.14.3 *Findings*

For the U.S. economy the mean annual growth rate in consumption was 0.018. Its standard deviation and autocorrelation were 0.036 and -0.14. Matching these facts necessitated setting $\mu = 0.018$, $\delta = 0.036$, and $\pi = 0.43$. Now, clearly the discount factor, β, should lie between 0 and 1. Mehra and Prescott (1985) suggest that the coefficent of relative risk aversion, α, is bounded between 0 and 10. So, they computed the risk-free rate and the equity premium for values of α and β that lie within these ranges subject to the condition that a solution for the model exists or that $|I - \beta\Lambda| \neq 0$. In other words, think about risk-free rate and the equity premium as being defined by two functions $R^f = R(\alpha, \beta)$ and $R^e - R^f = P(\alpha, \beta)$. They compute the values of these two functions for parameter values that lie in the set $\mathcal{X}$ where

$$\mathcal{X} = \{(\alpha, \beta) : 0 < \beta < 1, 0 < \alpha < 10, \text{ and } |I - \beta\Lambda| \neq 0\}.$$

As can be seen from Figure 8.14.1, the model can't simultaneously generate an equity premium of 6.98 percent and risk-free return of 0.8 percent. So, within the context of a frictionless Arrow-Debreu-McKenzie world it is difficult to rationalize why the average return on equity was so high while the risk-free return was so low.

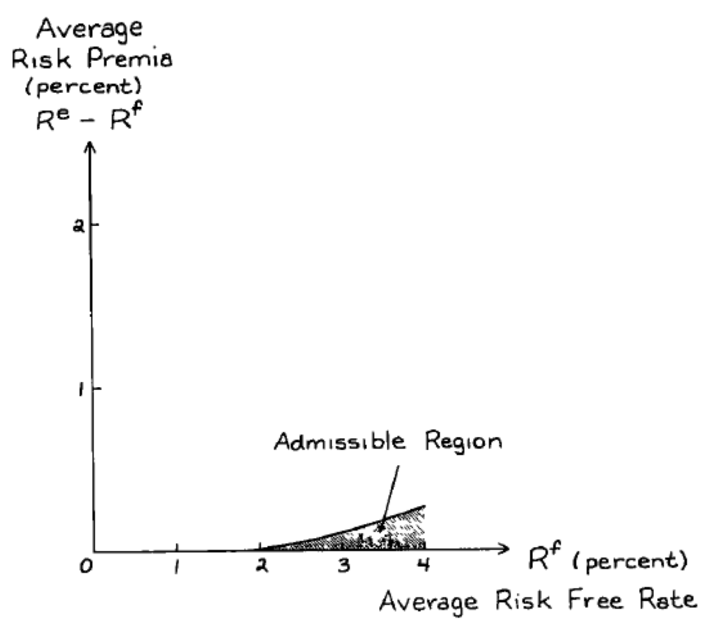


Figure 8.14.1: Equity premium and risk-free rate combinations for various values of the coefficient of relative risk aversion, α, and the discount factor, β, lying in the set $\mathcal{X}$. *Source*: Mehra and Prescott (1985, p. 155).

Kenneth J. Arrow (1921-), Gerard Debreu (1921-2004), and Lionel W. McKenzie (1919-2010) are considered to be the fathers of modern general equilibrium theory. Arrow and Debreu won Nobels in 1972 and 1983, respectively. In 1995 McKenzie was awarded The Order of the Rising Sun in Japan.

8.15 Approximating an AR1 by a Markov Chain

AR1 processes are commonly used in macroeconomics. The generic AR1 process has the form

$$z_{t+1} = \rho z_t + \varepsilon_{t+1}, \text{ where } \varepsilon_{t+1} \sim N(0, \sigma^2).$$

This has an autocorrelation coefficient of ρ, a conditional standard deviation of σ, and an unconditional (long-run) standard deviation of $\sigma/\sqrt{(1-\rho^2)}$.Here $0 < \rho < 1$ represents the coefficient of autocorrelation. Often the AR1 process is specified in logs. This ensures that z_{t+1} will always be nonnegative since now $z_{t+1} = z_t^{\rho} \exp(\varepsilon_{t+1})$. How can such AR1 processes be approximated by a N-state Markov chain, where $N \geq 2$.

8.15.1 Algorithm: Rouwenhorst (1995)

1. Constrain the variable z to always lie in a time-invariant grid of n *equally* spaced points centered around 0, so that $z \in \{z_1, \cdots, z_n\}$ with $-z_1 = z_n = \psi > 0$. where $\psi = \sigma\sqrt{n-1}/\sqrt{(1-\rho^2)}$. A transition matrix, T^n, is sought that has the form

$$T^n = \begin{bmatrix} \pi_{11} & \cdots & \pi_{1n} \\ \vdots & \ddots & \vdots \\ \pi_{n1} & \cdots & \pi_{nn} \end{bmatrix},$$

where π_{kl} are the odds of going *from* state k *to* state l. The summation across any row equals 1; i.e., $\sum_{l=1}^{n} \pi_{kl} = 1$ for all k.

2. The transition matrix T^j is generated recursively for $j = 3, \cdots, n$ as follows:

   (a)

$$\underset{j\times j}{T^j} = p \begin{bmatrix} \underset{(j-1)\times(j-1)}{T^{j-1}} & \underset{(j-1)\times 1}{\mathbf{0}} \\ \underset{1\times(j-1)}{\mathbf{0}'} & \underset{1\times 1}{0} \end{bmatrix} + (1-p) \begin{bmatrix} \underset{(j-1)\times 1}{\mathbf{0}} & \underset{(j-1)\times(j-1)}{T^{j-1}} \\ \underset{1\times 1}{0} & \underset{1\times(j-1)}{\mathbf{0}'} \end{bmatrix}$$
$$+ (1-p) \begin{bmatrix} \underset{1\times(j-1)}{\mathbf{0}'} & \underset{1\times 1}{0} \\ \underset{(j-1)\times(j-1)}{T^{j-1}} & \underset{(j-1)\times 1}{\mathbf{0}} \end{bmatrix} + p \begin{bmatrix} \underset{1\times 1}{0} & \underset{1\times(j-1)}{\mathbf{0}'} \\ \underset{(j-1)\times 1}{\mathbf{0}} & \underset{(j-1)\times(j-1)}{T^{j-1}} \end{bmatrix},$$

   where

$$T^2 = \begin{bmatrix} p & 1-p \\ 1-p & p \end{bmatrix},$$

   $\mathbf{0}$ is an $(j-1) \times 1$ *column* vector of zeros, and $p = (1+\rho)/2$.

   (b) At the end of each iteration, *all but* the first and last rows of T^j should be divided by 2.

The idea is that if you are in the upper left cell on iteration $j-1$ then you will stay there with probability p on iteration j or move to the upper right cell with the complementary probability $1-p$. Note that it is possible to get into the rows $2, \cdots, n-1$ of T^j from either the upper or lower cells, which explains the division by 2; i.e., without the division by 2, $\sum_l^n \pi_{kl} = 2$ for $k = 2, \cdots, n-1$. By setting $p = (1+\rho)/2$, the Markov chain will have the same long-run variance and first-order autocorrelation as the AR1.

8.16 *Interpolation*

Suppose that one wants to represent a set of n points, $(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$, by a continuous function $y = F(x)$, for which an analytical expression is not available. That is, assume that y_i represents the value of F when evaluated at the point x_i, for $i = 1, \cdots, n$. So, the problem is to construct a continuous function from the set of given data points, $(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$, that will specify a value for y given any value for x within some specified range, say for example $[x_1, x_n]$. The constructed function will have the property that $y_i = F(x_i)$ for all i, so that it fits all of the specified data points.

8.16.1 *Fitting a Polynomial*

A quick and dirty thing to do is to fit a polynomial through the points $(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$. That is, a m-th degree polynomial of the following form could be fit to data.

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \cdots + \beta_m x^m.$$

Now, in this case the resulting polynomial will *not* go through each of the points $(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$. When fitting a polynomial to data, it is best to keep the degree m as low as possible. The coefficients $\beta_0, \beta_1 \cdots, \beta_m$ are chosen to minimize the sum of the squared errors:

$$\min_{\beta_0, \beta_1 \cdots, \beta_n} \{[\sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i - \beta_2 x_i^2 - \cdots - \beta_m x_i^m)]^2\}.$$

The first-order conditions solving this problem are

$$\begin{array}{rcl}
\sum_{i=1}^n (\beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \cdots + \beta_m x_i^m) \times \sum_{i=1}^n 1 & = & \sum_{i=1}^n y_i \times \sum_{i=1}^n 1 \\
\sum_{i=1}^n (\beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \cdots + \beta_m x_i^m) \times \sum_{i=1}^n x_i & = & \sum_{i=1}^n y_i \times \sum_{i=1}^n x_i \\
\vdots & \vdots & \vdots \\
\sum_{i=1}^n (\beta_0 + \beta_1 x_i + \beta x_i^2 + \cdots - +\beta x_i^m) \times \sum_{i=1}^n x_i^m & = & \sum_{i=1}^n y_i \times \sum_{i=1}^n x_i^m
\end{array}.$$

This is a linear system in $\beta_0, \beta_1 \cdots, \beta_m$.[6]

In Figure 8.16.1 first- and second-order polynomials are fit to a time series for Chinese annual real GDP spanning the period 1952-2019.

[6] In matrix form this gives the familiar formula for an ordinary least squares regression: $\beta = (X'X)^{-1}X'y$.

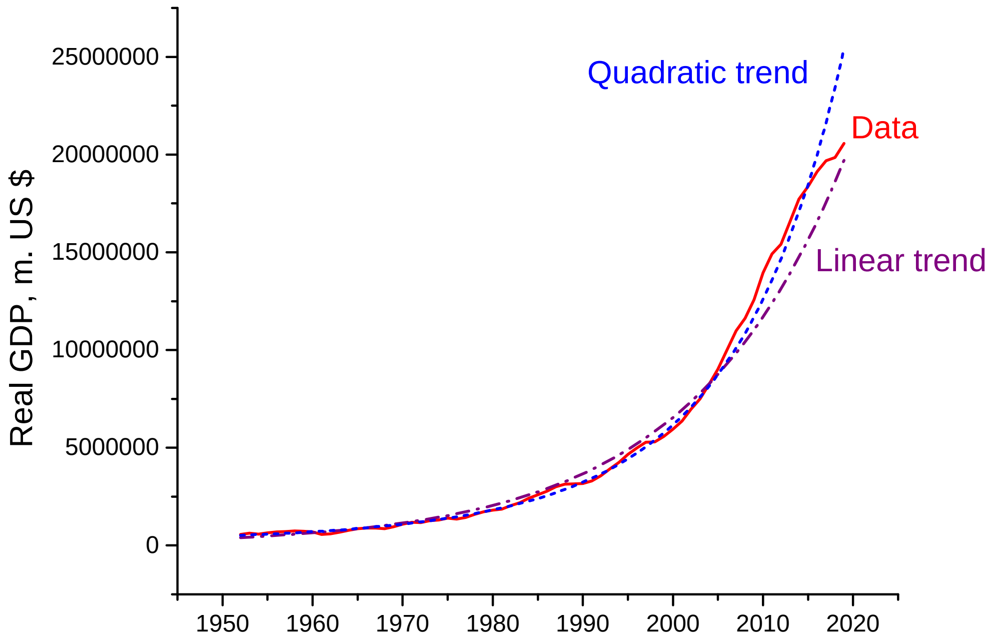


Figure 8.16.1: Chinese annual real GDP measured in millions of US dollars, 1952-2019. Two trend lines are fit through the series. The first uses a linear trend, the second a quadratic one. Over this time period China had an annual growth rate in real GDP of 5.4%.

This corresponds to running linear and quadratic time trends through the data. When fitting the polynomials, the time series for real GDP was first logged. After doing this, the predicted values from the logged data where then exponentiated (or unlogged) to express things again in millions of US dollars. Last, note that the polynomials do not go through all of the data points.

.

8.16.2 *Piecewise Linear Interpolation*

Piecewise linear interpolation places a continuous curve through the n points, $(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$. The segment between each pair of adjacent points is linear. So, the curve is made up of a bunch of linear segments. Denote this piecewise linear function by L. The piecewise linear function L satisfies the following criteria:

1. $L(x)$ is represented by a linear function on each of the subintervals $[x_h, x_{h+1}]$ for each $h = 1, ..., n-1$. Denote this linear function by

$$L^h(x) = \alpha_h + \beta_h(x - x_h), \text{ for } x_h \leq x \leq x_{h+1}.$$

2. $L(x) = y_h$ when $x = x_h$ for each $h = 1, ..., n-1$. So, the piecewise linear function passes through each interpolation point.

It is easy to see that following solution for $L^h(x)$ works:

$$L^h(x) = (1-\mu)y_h + \mu y_{h+1},$$

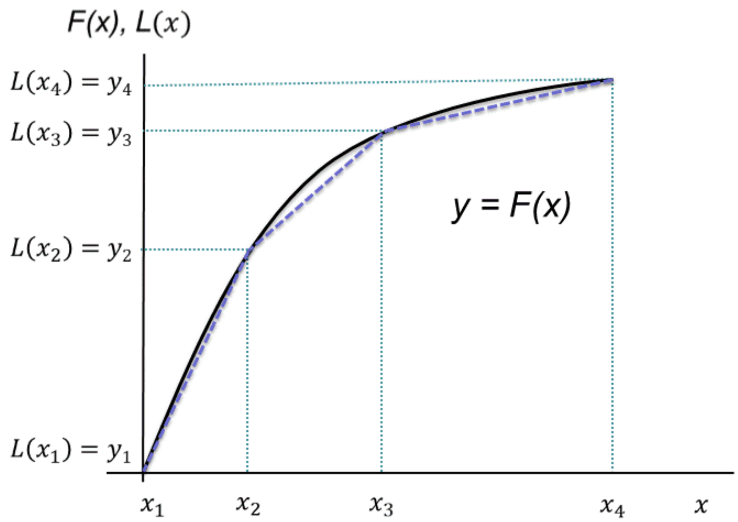


Figure 8.16.2: The function $y = F(x)$ is approximated by the piecewise linear function $L(x)$. At each interpolation point, x_i, the piecewise linear function, $L(x_i)$, takes the same value as $y_i = F(x_i)$. The values of $L(x)$ and $F(x)$ differ when not at an interpolation point.

$\cdots$ where $\mu = (x - x_h)/(x_{h+1} - x_h)$, for $x_h \leq x \leq x_{h+1}$. This implies

$$\beta_h = \frac{y_{h+1} - y_h}{x_{h+1} - x_h},$$

and

$$\alpha_h = y_h.$$

Figure 8.16.2 illustrates the situation, where the function $y = F(x)$ is approximated by the piecewise linear function $L(x)$.

8.16.3 *Cubic Spline Interpolation*

Cubic spline interpolation fits a flexible, C^2 curve through the n points, $(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$. Denote this spline by $S(x)$. The spline function S satisfies the following criteria:

1. $S(x)$ is made up by cubic polynomials, denoted by S^h, on each of the subintervals $[x_h, x_{h+1}]$ for each $h = 1, ..., n-1$. Denote this cubic by

   $$S^h(x) = \alpha_h + \beta_h(x - x_h) + \psi_h(x - x_h)^2 + \delta_h(x - x_h)^3, \text{ for } x_h \leq x \leq x_{h+1}.$$

   A prototypical cubic is given shown in Figure 8.16.3.

2. $S(x) = y_h$ when $x = x_h$ for each $h = 1, ..., n$. Therefore, it goes through all of the interpolation points.

3. $S^{h+1}(x_{h+1}) = S^h(x_{h+1})$ for each $h = 1, ..., n-2$. The cubics over each interval are connected.

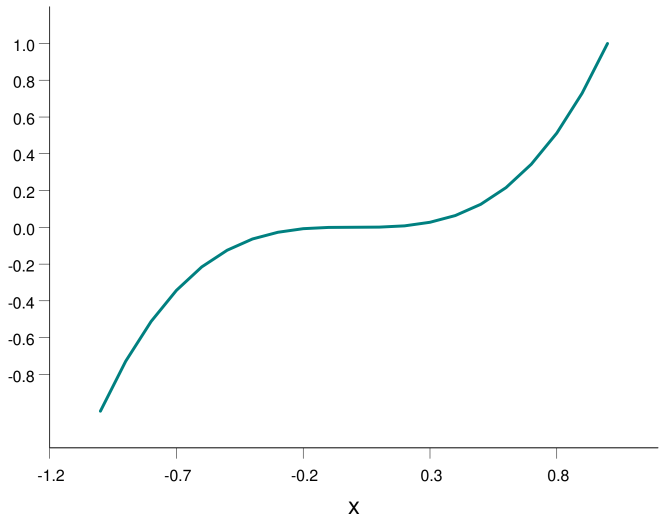


Figure 8.16.3: A Cubic Equation: $y = x^3$.

4. $S_1^{h+1}(x_{h+1}) = S_1^h(x_{h+1})$ for each $h = 1, ..., n-2$. The connections are smooth in the sense that the first derivatives are the same where one cubic ends and the other starts.

5. $S_{11}^{h+1}(x_{h+1}) = S_{11}^h(x_{h+1})$ for each $h = 1, ..., n-2$. The function is very smooth in the sense that the second derivatives are the same at connection points.

6. One of the following boundary conditions is satisfied:

   (a) $S_{11}(x_1) = S_{11}(x_n) = 0$ (free or natural),

   (b) $S_1(x_1) = F_1(x_1)$ and $S_1(x_n) = F_1(x_n)$ (clamped), where $F(x)$ is some function that is being approximated.

Observe that there are $n-1$ intervals. Hence, there are $4(n-1)$ parameters that a solution is need for–the α_h^j's, β_h^j's, ψ_h^j's and δ_h^j's. Properties 2 to 6 imply that there will be exactly $4(n-1)$ linear restrictions. The Hodrick-Prescott filter, discussed later, is a close cousin of cubic spline interpolation.

8.16.4 *Spline Art*

Cubic spline functions are very flexible and can be used to approximate many things. An artist's ink drawing of a face is shown in the upper panel of Figure 8.16.4. The lower panel shows a computer-generated facsimile of the artist's sketch using cubic spline functions. To do this, the picture on the left was broken up into 4 regions: the left eyebrow, the left eyeball, the profile, and the right eyelid. The pixel coordinates for the lines in each of the regions were read off by placing the cursor from a mouse on the parts of the lines in each region. The pixel coordinates were then translated into (x, y) coordinates. There are many programs that can read off pixel coordinates from a graph, such as Windows Paint. The more points that are read off, the more accurate will be the computer-generated rendition. Finally, 4 cubic splines are fit to (x, y) coordinates in each of the 4 regions. The cubic spline interpolation did a great job replicating the artist's sketch. This demonstrates the utility of cubic splines.

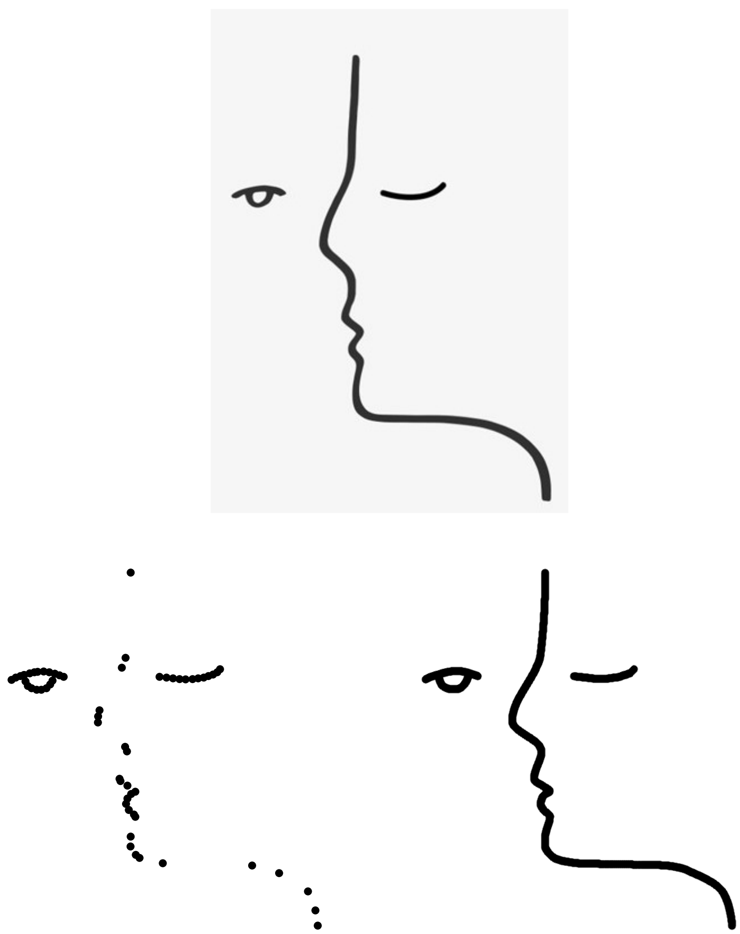


Figure 8.16.4: Spline Art. The top panel presents a sketch of face done by an artist. The left plot in the bottom panel shows a set of points mimicking the sketch. The right plot in the bottom panel is a computer generated approximation of the face using shape-preserving cubic spline functions.

8.16.5 *Radial Basis Functions*

A modern approach to interpolation is radial basis functions. An advantage of this approach is that is easily extends to multivariate functions. The idea underlying radial basis function interpolation is easy to understand. The interpolating function is a linear combination of radial basis functions each centered around one of the interpolation points, x_i, in the domain. The generic radial basis function has the form $\phi(|x - x_i|)$, where $|x - x_i|$ is the radial distance of the point x from the interpolation point x_i. Most often the Euclidian norm is used for the distance measure. Note that even though the x's could be vectors, ϕ is a function of only one argument; the radial distance, $r \equiv |x - x_i|$. An equisized step away from x_i in any direction has the same influence on ϕ. The value of the interpolating function at the point x is given by

$$R(x) = \sum_{i=1}^{n} \omega_i \phi(|x - x_i|),$$

where ω_i is the weight attached to the radial basis function that is centered at the point x_i. The value of the interpolating function at the point x is a function of all of the interpolation points, or the x_i's. To compute the weights, or the ω_i's, the function $R(x)$ is forced to have the value y_i when evaluated at x_i. Thus, the weights can be recovered by solving the following system of linear equations

$$\begin{array}{lll} y_1 & = R(x_1) = & \sum_{i=1}^{n} \omega_i \phi(|x_1 - x_i|) \\ \vdots & & \vdots \\ y_n & = R(x_n) = & \sum_{i=1}^{n} \omega_i \phi(|x_n - x_i|). \end{array}$$

Some examples of radial basis function are shown below (where ε is some constant):

VARIOUS RADIAL BASIS FUNCTIONS

$\phi(r) = e^{-\varepsilon r^2}$,	Gaussian
$\phi(r) = \sqrt{1 + (\varepsilon r^2)^2}$	Multiquadric
$\phi(r) = 1/\sqrt{1 + (\varepsilon r^2)^2}$	Inverse multiquadric
$\phi(r) = r^k$,for $k = 1, 3, 5, \cdots$	Polyharmonic spline, odd
$\phi(r) = r^{k-1} \ln(r^r)$, for $k = 2, 4, 6, \cdots$	Polyharmonic spline, even.

Figure 8.16.5 plots the Gaussian radial basis function for several values of ε. The value of the function ϕ declines as one moves away from the center, $r = |x - x_i| = 0$. Therefore, less value will be attached to points in the domain, x, further away from the point, x_i, that is being interpolated around. The shape parameter, ε, controls the speed of the decay. The bigger the value of ε, the less weight will assigned to distant points.

8.17 *The Hodrick-Prescott Filter*

The Hodrick and Prescott (1997) filter is often used in macroeconomics to detrend economic time-series . This filter draws a smooth curve (a cubic spline) through an economic time series. As such, it does not

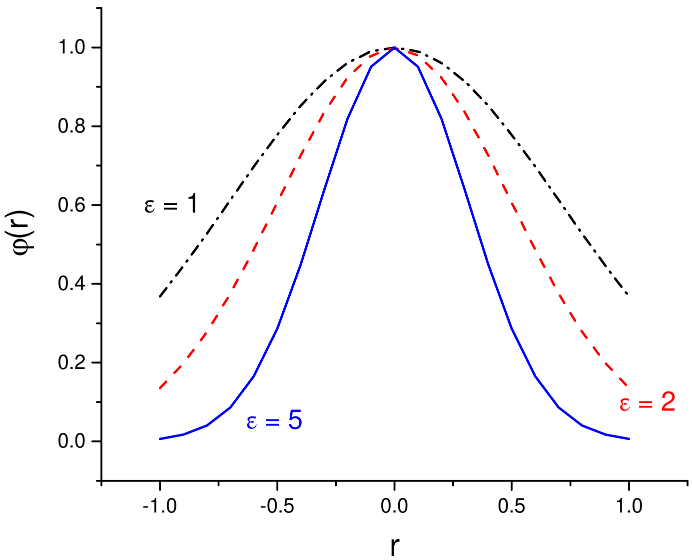


Figure 8.16.5: A Gaussian radial basis function plotted for three values of the shape parameter, ε.

pass through each data point. While controversial when introduced in 1981, it is actually a variant of the Whittaker (1923) cubic smoothing spline, which has a long and distinguished history in the statistics literature.[7] To see how it works, let $\{y_t\}_{t=1}^T$ represent some time series of interest. Usually, this time series has been logged. The filter fits a trend through series, denoted by $\{\tau_t\}_{t=1}^T$. This trend solves the following minimization problem:

$$\min_{\{\tau\}_{t=1}^T} \{\sum_{t=1}^T (y_t - \tau_t)^2 + \lambda \sum_{t=2}^{T-1} [(\tau_{t+1} - \tau_t) - (\tau_t - \tau_{t-1})]^2\}$$
$$= \min_{\{\tau\}_{t=1}^T} \{\sum_{t=1}^T (y_t - \tau_t)^2 + \lambda \sum_{t=2}^{T-1} (\tau_{t+1} - 2\tau_t + \tau_{t-1})^2\},$$

where λ is a constant that governs the degree of smoothness in the trend. As can be seen, changes in the first-differences of the trend are penalized. That is, "roughness" in the resulting curve is penalized. How much depends on the size of λ, the smoothing parameter.

1. When $\lambda = 0$ the minimization routine sets $y_t = \tau_t$, since no movements in the trend are penalized. One could think about fitting a cubic interpolating spline through the points $\{y_t\}_{t=0}^n$. This would set $\sum_{t=1}^T (y_t - \tau_t)^2 = 0$, by property 2 of the cubic interpolating spline. Thus, setting $\lambda = 0$ returns a cubic interpolating spline.

2. When $\lambda \to \infty$ the trend becomes linear since this will set the last term to zero. To see this, suppose that $\tau_t = a + bt$. Then, $\tau_{t+1} - \tau_t = b$ and $\tau_t - \tau_{t-1} = b$.

[7] While widely accepted now, the controversy explains the long delay in publication.

Sir Edmund Taylor Whittaker (1873-1956) was a well-known mathematician in his time. He was a professor of mathematics at the University of Edinburgh. In 1895 he graduated from the University of Cambridge as the Second Wranger in mathematics.

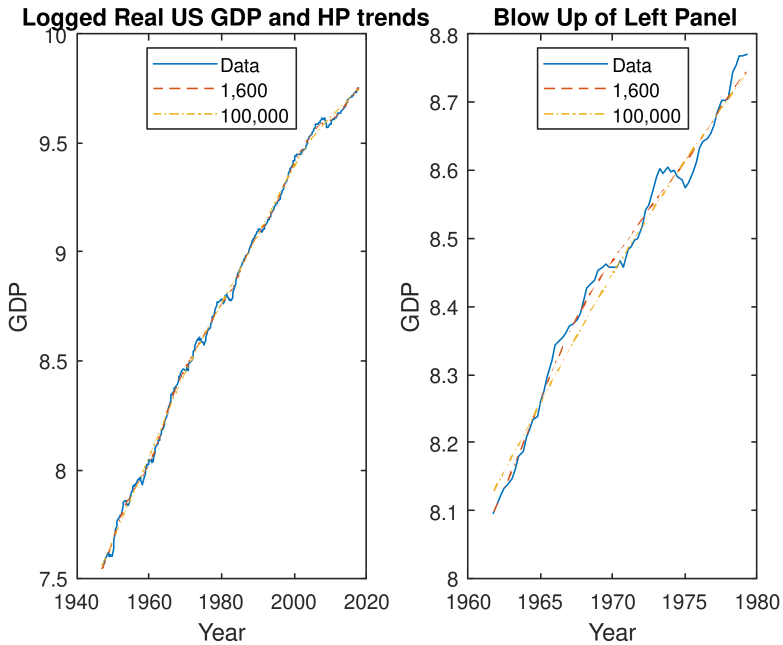


Figure 8.17.1: Quarterly real GDP and its HP trend, 1947-2017. Real GDP has been logged. The HP trend is shown for two values of the smoothing parameter, $\lambda = 1,600$ and $\lambda = 100,000$. As the smoothing parameter is increased the HP trend becomes less flexible. The right panel is merely a blow up of the left one.

3. For $0 < \lambda < \infty$ solve the above problem to get $\{\tau_t\}_{t=1}^T$. Observe that $\{\tau_t\}_{t=1}^T \neq \{y_t\}_{t=1}^T$, because roughness in the resulting curve is being penalized. The HP trend is obtained by fitting a cubic interpolating spline to the points $\{\tau_t\}_{t=1}^T$. One can think about this procedure as fitting a cubic interpolating spline, $S(t)$, to the data points $\{y_t\}_{t=1}^T$ while dropping the n restrictions that $S(t) = y_t$. These restrictions are made up by using the T first-order conditions to the above problem.

Often for quarterly data λ is set to 1,600. For annual data a value of 6.25 has been suggested (although values of 100 and 400 are also used). Figure 8.17.1 plots postwar quarter real GDP together with its H-P trend. H-P detrended is illustrated in Figure 8.17.2.

The generic first-order condition connected to the above minimization problem is

$$-(y_t - \tau_t) - \lambda 2(\tau_{t+1} - 2\tau_t + \tau_{t-1}) + \lambda(\tau_t - 2\tau_{t-1} + \tau_{t-2}) + \lambda(\tau_{t+2} - 2\tau_{t+1} + \tau_t) = 0, \tag{8.17.1}$$

for $t = 2, \cdots, T-2$. This can be rewritten as

$$\lambda\tau_{t-2} - 4\lambda\tau_{t-1} + (6\lambda + 1)\tau_t - 4\lambda\tau_{t+1} + \lambda\tau_{t+2} = y_t, \text{ for } \tau = 2, \cdots, T-2.$$

This first-order condition takes a more restricted form for $t = 1, 2$ and $t = T-1, T$. For example, it is easy to see that the terms in the objective function involving τ_1 are $(y_1 - \tau_1)^2 + \lambda(\tau_3 - 2\tau_2 + \tau_1)^2$. So the first-order condition for τ_1 is $-(y_1 - \tau_1) + \lambda(\tau_3 - 2\tau_2 + \tau_1) = 0$.

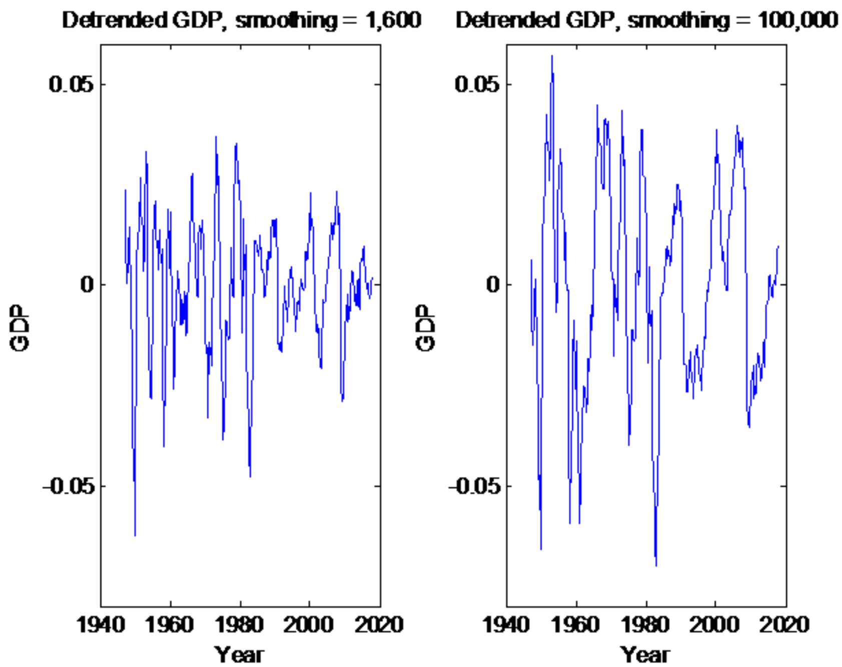


Figure 8.17.2: H-P detrended quarterly real GDP, 1947-2017. Real GDP was logged before it was detrended. The figure shows detrended GDP for two values of the smoothing parameter, namely $\lambda = 1,600$ and $\lambda = 100,000$. Observe how the fluctuations increase with the size of λ.

This can be expressed as

$$(\lambda + 1)\tau_1 - 2\lambda\tau_2 + \lambda\tau_3 = y_1.$$

Doing the same thing for τ_2, τ_{T-1}, and τ_T gives

$$-2\lambda\tau_1 + (5\lambda + 1)\tau_2 - 4\lambda\tau_3 + \lambda\tau_4 = y_2,$$

$$\lambda\tau_{T-3} - \lambda 4\tau_{T-2} + (5\lambda + 1)\tau_{T-1} - 2\lambda\tau_T = y_{T-1},$$

and

$$\lambda\tau_{T-2} - 2\lambda\tau_{T-1} + \lambda\tau_T = y_T.$$

This represents a system of T equations in the T unknowns, $\tau_1, \tau_2, \cdots, \tau_n$. This system is linear in the trend variables, $\tau_1, \tau_2, \cdots, \tau_n$, and hence can be solved using linear algebra.

To this end, construct the matrices shown below

$$T \equiv \begin{bmatrix} \lambda+1 & -2\lambda & \lambda & 0 & 0 & & 0 & 0 & 0 & 0 & 0 \\ -2\lambda & 5\lambda+1 & -4\lambda & \lambda & 0 & \cdots & 0 & 0 & 0 & 0 & 0 \\ \lambda & -4\lambda & 6\lambda+1 & 4\lambda & \lambda & & 0 & 0 & 0 & 0 & 0 \\ 0 & \lambda & -4\lambda & 6\lambda+1 & 4\lambda & & 0 & 0 & 0 & 0 & 0 \\ & & \vdots & & & & & & \vdots & & \\ 0 & 0 & 0 & 0 & 0 & & \lambda & -4\lambda & 6\lambda+1 & 4\lambda & \lambda \\ 0 & 0 & 0 & 0 & 0 & \cdots & 0 & \lambda & -4\lambda & 5\lambda+1 & -2\lambda \\ 0 & 0 & 0 & 0 & 0 & & 0 & 0 & \lambda & -2\lambda & \lambda+1 \end{bmatrix},$$

$$\tau = \begin{bmatrix} \tau_1 \\ \vdots \\ \tau_T \end{bmatrix}, \text{ and } y = \begin{bmatrix} y_1 \\ \vdots \\ y_T \end{bmatrix}.$$

Using these matrices the solution for the Hodrick-Prescott filter reads $T\tau = y$ or

$$\tau = T^{-1}y.$$

The detrended series is simply $y - \tau$ or $(I - T^{-1})y$.

9 Stochastic Dynamics

> "Our task as I see it ... is to write a Fortran program that will accept specific economic policy rules as 'input' and will generate as 'output' statistics describing the operating characteristics of time series we care about, which are predicted to result from these policies ... It must be taken for granted that simply attempting various policies that may be proposed on actual economies and watching the outcome must not be taken as a serious solution method: Social Experiments on the grand scale may be instructive and admirable, but they are best admired at a distance." (Robert E. Lucas Jr, "Methods and Problems in Business Cycle Theory," *Journal of Money,Credit and Banking*, 1980).

9.1 Introduction

The macroeconomy is full of randomness. For example, no one knows what the state of technology will be in the future. Think about how the information age is affecting the economy: artificial intelligence, robots in factories, online shopping, etc. Likewise, governments come and go, with different views about deficits, education, the environment, health care, income equality, and international trade. So, future spending and tax policies are unknown too. Acts of God, such as Covid19, earthquakes, and hurricanes, have effects too.

To capture this, let the state of the economy, defined by the vector (k_t, z_t), evolve according to

$$k_{t+1} = K(k_t, z_t),$$

where z_t is a random variable which is distributed according to the cumulative distribution function

$$z_{t+1} \sim G(z_{t+1}|z_t) = \Pr[\widetilde{z}_{t+1} \leq z_{t+1}|\widetilde{z}_t = z_t].$$

The associated density function is represented by $g(z_{t+1}|z_t) \equiv G_1(z_{t+1}|z_t)$. Observe that the randomness in the z's will imply randomness in the k's. Suppose that in period t one knows k_t and z_t. One will not know what z_{t+1} will be in period $t+1$, because this is random. This implies that $k_{t+2} = K(k_{t+1}, z_{t+1})$ will be unknown because it is a function of k_{t+1} and z_{t+1}. A long time ago, Eugen Slutsky (1937) discussed how the business cycle could be the result of random causes; this is discussed in Chapter 8.

The function K is modelled here as the outcome of a dynamic stochastic optimization problems by consumers and firms in conjunction with equilibrium conditions and the government budget constraint. This dynamic stochastic optimization problem is formulated as a dynamic programming problem à la Bellman (1957). Dynamic programming problems can be solved numerically in various ways. Three methods are covered here: discrete-state-space dynamic programming problem, linearization, and policy-function iteration. Three different algorithms for policy function iteration are discussed: the Coleman (1991) algorithm, the endogenous grid method by Carroll (2006), and parameterized expectations introduced by den Haan and Marcet (1990).

The above economy will never settle down to a deterministic steady state because z_t is always fluctuating. The best one can hope for is that in the long run fluctuations in (k_t, z_t) will be described by some stationary cumulative probability distribution, $S(k_t, z_t)$. The above two equations imply a joint probability distribution for (k_{t+1}, z_{t+1}) as a function of (k_t, z_t). Write this as

$$\mathbf{T}(k_{t+1}, z_{t+1}|k_t, z_t) \equiv \Pr[\widetilde{k}_{t+1} \leq k_{t+1}, \widetilde{z}_{t+1} \leq z_{t+1}|\widetilde{k}_t = k_t, \widetilde{z}_t = z_t], \tag{9.1.1}$$

where $\mathbf{T}$ is often referred to as the *transition operator*. The long-run distribution must solve the equation

$$S(k_{t+1}, z_{t+1}) = \int \int \mathbf{T}(k_{t+1}, z_{t+1}|k_t, z_t) dS(k_t, z_t).$$

Note that the righthand side of the above equation is counting the ways you can move into a situation in period $t+1$ where $\widetilde{k}_{t+1} \leq k_{t+1}$ and $\widetilde{z}_{t+1} \leq z_{t+1}$ from any of the possible (k_t, z_t) combinations in period t. In stochastic models one is interested in the statistical properties of (k_t, z_t), as opposed to characterizing a deterministic time path. This will be done two ways here. First, by using a Markov chain for $\mathbf{T}$ and, second, by simulating $\mathbf{T}$ using Monte Carlo techniques.

9.2 *Robinson Crusoe*

Consider the problem of Robinson Crusoe who lives on an island. Robinson must decide how much to consume and save in the form of capital each period. He does this to maximize the *expected* value of his lifetime utility

$$E[\sum_{t=1}^{\infty} \beta^{t-1} U(c_t)].$$

He produces according to the following production function

$$y_t = z_t F(k_t),$$

where z_t is a random technological shock in period t. In any given period t, output, y_t, is split between consumption, c_t, and gross investment, $k_{t+1} - (1-\delta)k_t$, where δ is the rate of depreciation on capital. Robinson makes his consumption and investment decision after he sees z_t. Let z_t follow a stochastic process of the following form

$$z_{t+1} \sim G(z_{t+1}|z_t) = \Pr[\widetilde{z}_{t+1} \leq z_{t+1}|\widetilde{z}_t = z_t].$$

Here G is the *cumulative* distribution function for z_{t+1} conditioned on z_t, with associated *density* function $g(z_{t+1}|z_t) \equiv G_1(z_{t+1}|z_t)$. The above setting is a simplified version of the famous Brock and Mirman (1972) stochastic growth model. As mentioned in Chapter 8, Slutsky (1937) viewed the business cycle as resulting from random perturbations to the economy. His analysis had more to do with statistical mechanics than with economics, however. The stochastic process for z_t is often operationalized using one of two forms. First, z_t can be assumed to follow an AR1 in logs. Specifying the process in logs ensures that the values drawn for z_t are always positive. Second, z_t can be taken to be governed by a Markov chain. Both of these forms for stochastic processes are discussed in Chapter 8. It's easy to add labor into the stochastic growth model along the lines outlined in Chapter 6.

William A. Brock (1941-) is a research professor in the Department of Economics at the University of Missouri. He started his career in 1969 as an Assistant Professor at the University of Rochester. There he collaborated with Leonard J. Mirman (1940-2017) to develop the stochastic growth model. Mirman was a graduate student at the time.

In any particular period, Robinson's world will determined by the capital stock he has, k, and the value of the technology shock, z. The pair (k, z) is called the 'state of the world.' Robinson will make all decisions based upon the state of his world. Let $V(k, z)$ denote the *maximal* expected lifetime utility that Robinson realizes if today's state of the world is (k, z). Robinson lives in a stationary world in the sense that everyday is the same as another except that he may have a different level of capital, k, and realize a different level of the technology shock, z. In this stationary environment, Robinson's decision problem is described by

$$\begin{aligned} V(k,z) &= \max_{k'}\{U\Big(\underbrace{zF(k) + (1-\delta)k - k'}\Big) + \beta \int V(k',z')dG(z'|z)\} \\ &= \max_{k'}\{U\Big(zF(k) + (1-\delta)k - k'\Big) + \beta \int V(k',z') \underbrace{g(z'|z)}_{\text{density for } G} dz'\}. \end{aligned} \tag{9.2.1}$$

Problem (9.2.1) is a dynamic programming problem. It is presented using recursive notation, where the time subscripts have been dropped in contrast to the determistic formulation covered in Chapter 6. In this problem effectively there is only today and tomorrow, where a $'$ is attached to variable to denote its value tomorrow. Today's utility is given by $U\Big(zF(k) + (1-\delta)k - k'\Big)$, whereas *expected* discounted lifetime utility from tomorrow on is represented by $\beta \int V(k', z')g(z'|z)dz'$.

As with the deterministic version of the neoclassical growth model, it can be shown the that the value function, V, exists and is unique, is increasing, and strictly concave in k, and is continuously differentiable in k–the analysis proceeds along the lines of Section 6.8 in Chapter 6.[1]

[1] The only additional assumption needed is that $\int V(k',z')dG(z'|z)$ is continuous in k', if $V(k',z')$ is. This is known as the Feller property. See Stokey and Lucas (1986) for the formalities.

In problem (9.2.1) Robinson uses the distribution function $G(z'|z)$ to forecast future shocks, z', contingent on the current technology shock, z. Therefore, the probability distribution that Robinson uses for forecasting z' coincides with the actual probability distribution governing the evolution of z'; that is, Robinson's *subjective* beliefs about the evolution of z' are the same as the *objective* probability distribution for z' so that his expectations are *rational*.

Robert E. Lucas, Jr (1937-2023) was instrumental in bringing forth the notation of *rational expectations* into macroeconomics. This is now standard operating procedure for macroeconomists. Departing from this assumption is problematic. The issue is how to pin people's subjective expectations when they do not coincide with the objective expectations. In models of risky behavior, such as recreational drug use, it may be reasonable to assume that people do not know the true odds of addiction or death by overdose. But, reliable evidence on such things is often scant. An early example of the rational expectations hypothesis is the classic by Lucas and Prescott (1971) titled "Investment under Uncertainty."

Let the solution for k' that arises out of this problem be represented by

$$k' = K(k,z). \tag{9.2.2}$$

This is called Robinson's 'decision rule. It gives Robinson's optimal action for capital accumulation should he find himself in the state of the world (k,z). It is defined for every (k,z) pair that could occur on the island. Solving the above problem leads to the following first-order condition:

$$\underbrace{U_1\Big(zF(k) + (1-\delta)k - k'\Big)}_{\text{MC of invest}} = \underbrace{\beta \int V_1(k',z')dG(z'|z)}_{\text{MB of invest}} \tag{9.2.3}$$
$$= \beta \int V_1(k',z')g(z'|z)dz'.$$

The lefthand side of the above expression is the marginal cost of investing an extra unit of capital today. The righthand side is the discounted expected benefit. To understand this, note that $V_1(k',z')$ is the expected benefit next period from having an extra unit of capital should the technology shock be z'. But, the value of next period's technology shock is unknown currently so Crusoe must take the expected value of this benefit. Observe that the current value of the technology shock, z, is useful for forecasting the value of z' through the cumulative distribution function $G(z'|z)$ [or equivalently through its associated probability density function, $g(z'|z)$]. Last, since this expected benefit occurs next period it must be discounted by β. This first-order condition represents one equation in the one unknown, k', where k and z are known exogenous variables in the current period. Hence, the solution for k' will have the form shown by (9.2.2).

If the technology shock is a discrete random variable taking n values, so that $z \in \mathcal{Z} = \{z_1, z_2, \cdots, z_n\}$, then the dynamic programming problem (9.2.1) appears as

$$V(k,z_r) = \max_{k'}\{U\Big(z_rF(k) + (1-\delta)k - k'\Big) + \beta \sum_{s=1}^{n} \pi_{rs} V(k',z_s)\}, \tag{9.2.4}$$

where π_{rs} represents the odds of z traveling from z_r today to z_s tomorrow. Here the technology shock is described by an n-state Markov chain–see Chapter 8.

9.2.1 The Envelope Theorem, Again

This first-order condition involves the derivative of the unknown function V. It can be gotten rid off in the same way as in Chapter 6. To do this, let $\widetilde{k}'$ denote the *optimal* level of investment. Then, (9.2.1) can be written as

$$V(k,z) = U\Big(zF(k) + (1-\delta)k - \widetilde{k}'\Big) + \beta \int V(\widetilde{k}', z')dG(z'|z).$$

To eliminate $V_1(k', z')$ in (9.2.3), differentiate both sides of the above equation with respect to k. One gets

$$\begin{aligned} V_1(k,z) = {} & U_1\Big(zF(k) + (1-\delta)k - \widetilde{k}'\Big) \times [zF_1(k) + 1 - \delta] \\ & \underbrace{-U_1\Big(zF(k) + (1-\delta)k - \widetilde{k}'\Big)\frac{d\widetilde{k}'}{dk} + \beta \int V_1(\widetilde{k}', z')dG(z'|z)\frac{d\widetilde{k}'}{dk}}_{=0}, \end{aligned}$$

where the term on the second line is zero from (9.2.3). The fact that perturbations in the choice variable, k', cancel out in the objective function when evaluated at the optimal solution is called the envelope theorem.

9.2.2 The Stochastic Euler Equation

Updating the above result gives

$$V_1(k', z') = U_1\Big(z'F(k') + (1-\delta)k' - k''\Big) \times [z'F_1(k') + 1 - \delta].$$

Using this in the first-order condition (9.2.3) then leads to the following stochastic Euler equation:

$$\begin{aligned} U_1(zF(k) + (1-\delta)k - k') &= \beta \int U_1\Big(z'F(k') + (1-\delta)k' - k''\Big) \times [z'F_1(k') + 1 - \delta]dG(z'|z) \\ &= \beta E\Big[U_1\Big(z'F(k') + (1-\delta)k' - k''\Big) \times [z'F_1(k') + 1 - \delta]|k, z\Big]. \end{aligned} \tag{9.2.5}$$

9.2.3 Sequence Space Formulation

Robinson Crusoe's problem can also be cast in sequence space. Let the technology shock, z, follow an N-state Markov chain, with the transition probability between states i and j being denoted by π^{ij}, where the states are now represented by superscripts. Suppose that the technology shock starts off in period 1 from the known state i, so that

$z_1 = z_1^i$, where a superscript represents a state and a subscript the time period. Let $\mathbf{z}_t = (z_t, z_{t-1}, \cdots, z_1)$ signify some realized sequence of technology shocks between periods 1 and t. There are N^{t-1} possible sequences, with the set of these sequences notated by $\mathcal{Z}_t$. Denote the odds of the sequence $\mathbf{z}_t$ occurring by $\rho(\mathbf{z}_t)$. The probabilities $\rho(\mathbf{z}_t)$ are given recursively by

$$\begin{aligned}
\rho(\mathbf{z}_1^i) &= 1, \\
\rho[\mathbf{z}_2 = (z_2, \mathbf{z}_1^i)] &= \Pr(z_2 \mid \mathbf{z}_1)\rho(\mathbf{z}_1^i), \\
\rho[\mathbf{z}_3 = (z_3, \mathbf{z}_2)] &= \Pr(z_3 \mid \mathbf{z}_2)\rho(\mathbf{z}_2), \\
&\vdots \\
\rho[\mathbf{z}_t = (z_t, \mathbf{z}_{t-1})] &= \Pr(z_t \mid \mathbf{z}_{t-1})\rho(\mathbf{z}_{t-1}).
\end{aligned}$$

Think about Robinson Crusoe choosing a capital stock for every state-time combination that can possibly occur. That is, for each period t Robinson Crusoe chooses the capital stock for the next period, k_{t+1}, contingent upon the sequence of technology shocks that have occurred up to that period, or $\mathbf{z}_t$. This choice variable is designated by $k_{t+1}(\mathbf{z}_t)$. Note the capital stock chosen for period $t+1$ will depend on the initial capital stock, k_1, and the sequence of shocks that transpires between period 1 and period t, or $(z_t, z_{t-1}, \cdots, z_1)$.

Robinson Crusoe's maximization problem now appears as

$$\max_{\{k_{t+1}(\mathbf{z}_t)\}_{t=1}^{\infty}} \left\{ \sum_{t=1}^{\infty} \sum_{\mathbf{z}_t \in \mathcal{Z}_t} \beta^{t-1} \rho(\mathbf{z}_t) U\left(F(k_t(\mathbf{z}_{t-1})) + (1-\delta)k_t(\mathbf{z}_{t-1}) - k_{t+1}(\mathbf{z}_t)\right) \right\}.$$

Note that for each value of $\mathbf{z}_t$, the choice variable $k_{t+1}(\mathbf{z}_t)$ appears in just two periods in the maximand, namely time t and $t+1$. Specifically, it shows up in the terms

$$\begin{aligned}
&\cdots + \beta^{t-1}\rho(\mathbf{z}_t)U\left(z_t F(k_t(\mathbf{z}_{t-1})) + (1-\delta)k_{t+1}(\mathbf{z}_{t-1}) - k_{t+1}(\mathbf{z}_t)\right) \\
&+ \beta^t \sum_{j=1}^{N} \rho\left((z_{t+1}^j, \mathbf{z}_t)\right) U\left(z_{t+1}F(k_{t+1}(\mathbf{z}_t)) + (1-\delta)k_{t+1}(\mathbf{z}_t) - k_{t+2}(\mathbf{z}_{t+1})\right) + \cdots .
\end{aligned}$$

By writing $\rho\left((z_{t+1}^j, \mathbf{z}_t)\right) = \Pr(z_{t+1}^j \mid \mathbf{z}_t)\rho(\mathbf{z}_t)$ the above can be expressed as

$$\begin{aligned}
&\cdots + \beta^{t-1}\rho(\mathbf{z}_t)\{U\left(z_t F(k_t(\mathbf{z}_{t-1})) + (1-\delta)k_{t+1}(\mathbf{z}_{t-1}) - k_{t+1}(\mathbf{z}_t)\right) \\
&+ \beta^t \sum_{j=1}^{N} \Pr(z_{t+1}^j \mid \mathbf{z}_t) U\left(z_{t+1}F(k_{t+1}(\mathbf{z}_t)) + (1-\delta)k_{t+1}(\mathbf{z}_t) - k_{t+2}(\mathbf{z}_{t+1})\right)\} + \cdots .
\end{aligned}$$

Now, suppose that the current value of the technology shock in period t is z_t^i so that $\mathbf{z}_t$=$(z_t^i, \mathbf{z}_{t-1})$. Then, $\Pr(z_{t+1}^j \mid \mathbf{z}_t^i) = \pi^{ij}$ and the generic first-order condition reads

$$U_1\left(z_t^i F(k_t(\mathbf{z}_{t-1})) + (1-\delta)k_{t+1}(\mathbf{z}_{t-1}) - k_{t+1}(\mathbf{z}_t)\right)$$
$$= \beta \sum_{j=1}^{N} \pi^{ij} U_1\left(z_{t+1}^j F(k_{t+1}(\mathbf{z}_t)) + (1-\delta)k_{t+1}(\mathbf{z}_t) - k_{t+2}(\mathbf{z}_{t+1})\right)$$
$$\times [z_{t+1}^j F_1(k_{t+1}(\mathbf{z}_t)) + (1-\delta)],$$

for $i = 1, 2, \cdots, N$. With a minor switch to recursive notation, this can be rewritten as

$$U_1(zF(k) + (1-\delta)k - k')$$
$$= \beta E\left[U_1(z'F(k') + (1-\delta)k' - k'') \times [z'F(k') + (1-\delta)] \mid k, z\right].$$

This is the same stochastic Euler equation that obtained from the dynamic programming formulation.

9.3 *Business Cycle Modeling*

The goal of a business cycle model is to reproduce a set of stylized facts characterizing business cycles. This research program was laid out by Kydland and Prescott (1982) in one of the 20th century's most influential papers in macroeconomics. To characterize the business cycle, the U.S. time series for variables that one cares about, say GDP, consumption, investment, and hours worked, are first logged and then detrended using some filtering technique. Volatility of the logged and detrended series is measured by its standard deviation. (See Chapter A for a discussion of standard deviations, correlations, and autocorrelations.) To determine the cyclicality of a series its correlation with output is computed. A series is called procyclical when the correlation is positive. Last, persistence is judged by a series' autocorrelation. Kydland and Prescott (1982) matched up the predictions from their business cycle model with such a set of stylized facts for the U.S. economy. The original Kydland and Prescott (1982) paper picked the standard deviation of the technological shock, σ, and its autocorrelation, ρ, so that the model could match the observed standard deviation of output and its autocorrelation in the U.S. data.

Finn E. Kydland (1943-) is a Norwegian macroeconomist. He studied under Edward C. Prescott at Carnegie Mellon University. Along with Prescott, he won the Nobel Prize in 2004. Kydland relays that as an undergraduate at the Norwegian School of Economics and Business Administration he took a course that covered Ronald A. Howard's book on *Dynamic Programming and Markov Processes*. As an undergraduate, he wrote his first computer program doing dynamic programming in FORTRAN.

9.3.1 *Stylized Facts*

1. *Volatility*: Volatility is measured by standard deviation of the detrended logged variable. Investment is much more volatile than output, consumption less.

2. *Correlations*: Here the correlation between the detrended logged variable and detrended logged output are computed. Hours has

the highest correlation with output, but the other variables particularly consumption come fairly close.

3. *Persistence*: Now, the correlation between the detrended logged variable and its own lagged value is computed. In the data, consumption and productivity have the highest autocorrelations, and investment the lowest.

Business Cycle Statistics – Annual U.S. Data

Variable–logged	*Standard Deviation*	*Correlation*	*Autocorrelation*
Output	3.5	1.00	0.66
Consumption	2.2	0.74	0.72
Investment	10.5	0.68	0.25
Hours Worked	2.1	0.81	0.39
Productivity	2.2	0.82	0.77

The logged variable were detrended using a linear-quadratic time trend. *Source*: Greenwood et al. (1988)

9.4 Discrete-State-Space Dynamic Programming

This method has two key steps. In the first step, a dynamic programming problem is solved assuming that values for the capital stock and the technology shock both lie in discrete sets. In the second step the solution to this dynamics programming problem is represented as Markov chain. Using this Markov chain representation, statistics for any variable of interest can then be readily computed.

The capital stock in each period is constrained to be an element of the finite time-invariant set, $\mathcal{K}$. Thus,

$$k \in \mathcal{K} = \{k_1, \cdots, k_n\}.$$

For simplicity let the technology shock follow a two-state Markov chain so that

$$z \in \mathcal{Z} = \{z_1, z_2\},$$

with the transition probabilities

$$\pi_{rs} = \Pr[z' = z_s | z = z_r].$$

9.4.1 Representative Agent's Dynamic Programming Problem

Given the structure outlined above, the representative agent's dynamic programming problem can be written as

$$V(k_i, z_r) = \max_{c>0, k' \in \mathcal{K}} \{U(z_r F(k_i) + (1-\delta)k_i - k') + \beta \sum_{s=1}^{2} \pi_{rs} V(k', z_s)\}. \quad \text{P(1)}$$

Observe that $V : \mathcal{K} \times \mathcal{Z} \to \mathcal{R}$ is merely a list of $2n$ values, one for each $(k_i, z_r) \in \mathcal{K} \times \mathcal{Z}$. The choice for k' is restricted to lie in the discrete set $\mathcal{K}$. Discrete maximization was introduced in Chapter 3.

So, how can a solution V be obtained? Here's an algorithm.

1. Make an initial guess for V, denoted by V^0. This guess is merely a list of $2n$ values. Go to Step 2.

2. Enter iteration $j+1$ with a guess for the V on the righthand side of P(1). Call this guess V^j, it's merely a list of $2n$ values. Next, compute the solution to the righthand side of P(1). Denote this solution by V^{j+1}. To obtain it, let

$$M^j(k_i, z_r, k') = \{U(z_r F(k_i) + (1-\delta)k_i - k') + \beta \sum_{s=1}^{2} \pi_{rs} V^j(k', z_s)\}.$$

   This is the value of objective function in state (k_i, z_r) assuming that the capital stock $k' \in \mathcal{K}$ is chosen. In general, it will not be optimal to choose k'. It's easy to see that V^{j+1} is given by

$$\begin{aligned}
V^{j+1}(k_1, z_1) &= \max \overbrace{\{M^j(k_1, z_1, k_1), M^j(k_1, z_1, k_2), \cdots, M^j(k_1, z_1, k_n)\}}^{\text{n elements in set}}, \\
V^{j+1}(k_2, z_1) &= \max\{M^j(k_2, z_1, k_1), M^j(k_2, z_1, k_2), \cdots, M^j(k_2, z_1, k_n)\}, \\
&\vdots \\
V^{j+1}(k_n, z_1) &= \max\{M^j(k_n, z_1, k_1), M^j(k_n, z_1, k_2), \cdots, M^j(k_n, z_1, k_n)\}, \\
V^{j+1}(k_1, z_2) &= \max\{M^j(k_1, z_2, k_1), M^j(k_1, z_2, k_2), \cdots, M^j(k_1, z_2, k_n)\}, \\
&\vdots \\
V^{j+1}(k_n, z_2) &= \max\{M^j(k_n, z_2, k_1), M^j(k_n, z_2, k_2), \cdots, M^j(k_n, z_2, k_n)\}.
\end{aligned}$$

   This constitutes a revised guess for V. For each possible current state (k_i, z_r) the value for the future capital stock, k', that maximizes the objective function, say k_l, is found; that is, $M^j(k_i, z_r, k_l) \geq M^j(k_i, z_r, k_m)$, for all $m \neq l$. Note that max is a built in operation in MATLAB–see Appendix B. Go to Step 3.

3. Check whether $|V^{j+1} - V^j|$ is sufficiently small. If so, stop. If not, go back to Step 2. Essentially, equation P(1) defines an operator T such that $V^{j+1} = TV^j$. The contraction mapping theorem–see Chapter 6–states that the initial guess, V^0, is irrelevant, the solution for V is unique, and the algorithm will converge from V^0 to V. You could set $V^0 = 0$.

Decision Rule for Capital

The decision rule for capital, $K : \mathcal{K} \times \mathcal{Z} \to \mathcal{K}$, is

$$k' = K(k_i, z_r) \in \mathcal{K}.$$

This gives an investment plan for all $2n$ contingencies in the state space. All of the model's variables, such as consumption, gross investment, and output, can be written as a function of the current state of the world, (k_i, z_r):

$$o = O(k_i, z_r) = z_r F(k_i),$$

$$i = I(k_i, z_r) = K(k_i, z_r) - (1-\delta)k_i,$$

and

$$c = C(k_i, z_r) = z_r F(k_i) + (1-\delta)k_i - K(k_i, z_r).$$

It is easy to add labor, h, into the stochastic growth model along the lines presented in Chapter 6. Now the production function will read $y = zF(k,h)$. The decision rule for labor would have the form $h = H(k_i, z_r)$. Therefore, output in the current period can be expressed as $o = O(k_i, z_r) = z_r F\big(k, H(k_i, z_r)\big)$.

9.4.2 *Casting the Model's Solution as a Markov Chain*

The solution to the above model will now be cast as a Markov chain. The concept of a Markov chain was introduced in Chapter 8. The decision rule for capital can be rewritten in probabilistic form as

$$\Pr[k' = k_j \mid k = k_i, z = z_r] = \begin{cases} 1, & \text{for some } j, \\ 0, & \text{for the rest.} \end{cases}$$

Trivially, then

$$\sum_{j=1}^{n} \Pr[k' = k_j \mid k = k_i, z = z_r] = 1 \text{ for all } (k,z) \in \mathcal{K} \times \mathcal{Z}.$$

Define the transition probability between (k,z) pairs by

$$p_{ir,js} = \Pr[k' = k_j, z' = z_s | k = k_i, z = z_r] = \Pr[k' = k_j \mid k = k_i, z = z_r]\pi_{rs}. \tag{9.4.1}$$

Now, load these transition probabilities into a matrix:

$$T = \underbrace{[p_{ir,js}]}_{2n \times 2n}.$$

The matrix T is the Markov chain analogue to the transition operator described by equation (9.1.1). Observe that $\sum_{j,s} p_{ir,js} = 1$ for all i, r; i.e., from any state (i,r) you must go somewhere. Thus, each *row* in T sums to one.

Given some initial $1 \times 2n$ probability distribution ρ^0 over the state space $\mathcal{K} \times \mathcal{Z}$, next period's probability distribution is given by

$$\underset{1\times 2n}{\rho^1} = \underset{1\times 2n}{\rho^0} \times \underset{2n\times 2n}{T}.$$

The m-period-ahead probability distribution over state space $\mathcal{K} \times \mathcal{Z}$ reads

$$\rho^m = \rho^{m-1}T = \rho^{m-2}T^2 = \cdots = \rho^0 T^m.$$

The long-run or stationary distribution, ρ^*, solves

$$\rho^* = \rho^* T, \tag{9.4.2}$$

The stationary distribution can be computed using one of the methods discussed in Chapter 8.

Computation of Moments

Once the long-run distribution, ρ^*, has been obtained, it is easy to compute any moment of interest.

$$E[\ln o] = \sum_{r=1}^{2}\sum_{i=1}^{n}\rho^*_{ir} \ln O(k_i, z_r), \tag{9.4.3}$$

$$E[\ln c \ln o] = \sum_{r=1}^{2}\sum_{i=1}^{n}\rho^*_{ir} \ln C(k_i, z_r) \ln O(k_i,z_r), \tag{9.4.4}$$

$$E[\ln o' \ln o] = \sum_{s=1}^{2}\sum_{j=1}^{n}\sum_{r=1}^{2}\sum_{i=1}^{n} p_{ir,js}\rho^*_{ir} \ln O(k'_j, z'_s) \ln O(k_i, z_r). \tag{9.4.5}$$

To compute the percentage standard deviation of output use the formula

$$\sigma_{\ln o} = \sqrt{E[(\ln o)^2] - E[(\ln o)]^2}. \tag{9.4.6}$$

Similarly, the correlation between consumption and output is

$$\rho_{\ln c,\ \ln o} = \frac{E[\ln c \ln o] - E[\ln c]E[\ln o]}{\sigma_{\ln c}\sigma_{\ln o}}. \tag{9.4.7}$$

Last, the autocorrelation of output can be written as

$$\rho_{\ln o',\ \ln o} = \frac{E[\ln o' \ln o] - E[\ln o]^2}{(\sigma_{\ln o})^2}. \tag{9.4.8}$$

Choosing the Grid for the Capital Stock

How should the grid for the capital stock be chosen? One way of doing this is to plot the marginal distribution for the capital stock. The marginal distribution for capital is given by $(\rho^*_{11} + \rho^*_{12}, \rho^*_{21} + \rho^*_{22}, \cdots, \rho^*_{n1} + \rho^*_{n2})$. That is, the marginal distribution for capital is constructed by taking the joint distribution over capital and technology shocks and summing (or integrating) over the technology shock. It should be centered around the deterministic steady-state level of the capital stock, k^*. When the grid is picked correctly this distribution will resemble a choppy normal distribution. The odds of getting a very low or high capital stock (as given by the tails of distribution) will be small. Sometimes it helps to visualize the marginal distribution using a bar chart.

9.4.3 Algorithm: Discrete-State-Space Dynamic Programming

To summarize, the key steps in discrete-state-space dynamic programming are:

1. Solve the discrete-state-space dynamic programming problem P(1) to obtain the nonlinear decision rule for capital. Note that for a symmetic two-state Markov chain, the long-run variance and persistence of the technology shock map into unique values for $-z_1 = z_2 \equiv z$ and $\pi_{11} = \pi_{22} \equiv \pi$, as discussed in Chapter 8. So, one can think about setting the long-run variance and autocorrelation for the technology shock instead of setting values for the z's and π's. When solving the dynamic programming problem, it is important to ensure that consumption, c, is always positive. This can be done by putting a line such as the following into the code: $c = \max\{1.0E-8, c\}$. Likewise, when included, a lower bound may have to be imposed on labor supply.

2. Use the decision rule for capital and the Markov process for the technology shock to specify the Markov transition matrix (9.4.1).

3. Compute the stationary distribution for the capital stock and technology shock defined by equation (9.4.2) using one of the methods discussed in Chapter 8.

4. Compute various business cycle statistics using formulas such as (9.4.3) to (9.4.8).

Alternatively use the Monte Carlo based algorithm discussed below. This is less desirable in general. But in some circumstances one might want to filter the business cycle data coming from the model or perhaps the Markov transition matrix (9.4.1) is too big to handle.

1. Solve the discrete-state-space dynamic programming problem P(1) to obtain the nonlinear decision rule for capital. Again, when solving the dynamic programming problem, it is important to ensure that consumption, c, is always positive. This can be done by putting a line such as the following into the code: $c = \max\{1.0E-8, c\}$. Likewise, when included, a lower bound may have to be imposed on labor supply.

2. Draw a sample of T random variables, $\{\varepsilon_t\}_{t=1}^{T}$–random number generation is discussed in Chapter 8. In MATLAB a sample of uniformly distributed random variables on the $[0, 1]$ interval can be drawn using the RAND command. Make sure that the seed is fixed for the random number generator–this is done with the RNG(SEED), where SEED is some natural number. This line should be inserted just before the call for the random numbers.

3. Enter period t with a level of capital, $k_t = k_i \in \mathcal{K}$, and some past value for the technology shock, $z_{t-1} = z_r \in \mathcal{Z}$. The current technology shock could remain at its past value, z_r, or switch to a new value, z_s. Now, take ε_t from the sample of random variables. Compute the current technology shock, z_t, for its symmetric two-state Markov chain as follows:

$$z_t = z_r, \text{ if } \varepsilon < \pi_{rr},$$
$$z_t = z_s, \text{ if } \varepsilon \geq \pi_{rr}.$$

   Compute next period's capital stock, $k_{t+1} \in \mathcal{K}$, using the decision rule for capital
$$k_{t+1} = K(k_t, z_t).$$

4. For the starting value of capital just take $k_1 = k^*$, where k^* is level of capital in the deterministic steady state. For z_0 take either value for the technology shock.

5. Given a sample path of capital stocks and technology shocks, $\{k_t\}_{t=1}^{T+1}$ and $\{z_t\}_{t=1}^{T+1}$, data for all other variables of interest in the model, say consumption, investment and GDP, can be calculated. Sometimes researchers throw away some numbers at the beginning, say $\{k_t\}_{t=1}^{n}$. This way the sample is not influenced by the starting values for k_1 and z_0. From these variables one can then calculate a set of business cycle statistics. When doing this take the logarithm of variable. In MATLAB standard deviations can be computed using the STD command. Likewise, correlations can be calculated using the CORRCOEF command. The stationary distribution for any variable can be displayed using a histogram.

9.5 *Linearization*

This method involves taking a log-linear approximation of the Euler equation (9.2.5). The algorithm involves three steps: (i) conjecturing a log-linear law of the motion for capital accumulation; (ii) log-linearizing the Euler equation, while utilizing this guess, and solving for the resulting log-linear law of motion for capital while imposing a consistency requirement between the conjectured decision rule and the log-linearized solution; (iii) undertaking a Monte Carlo simulation of the computed decision rule to obtain sample paths for the variables of interest; and (iv) then computing the business facts that result from these sample paths.

9.5.1 Conjecturing a Decision Rule

Suppose that the technology shock follows an AR1 form specified by

$$\ln z' = \rho \ln z + \varepsilon', \text{ where } \varepsilon' \sim N(0, \sigma^2). \tag{9.5.1}$$

Conjecture a log-linear decision rule for capital of the following form

$$\ln k' = a + b \ln k + f \ln z. \tag{9.5.2}$$

One would expect that $a > 0$, $0 < b < 1$, and $f > 0$. To solve for a, b, and f, the above stochastic Euler equation (9.2.5) will linearized in $\ln k$ and $\ln z$. The resulting log-linear solution must be consistent with the assumed one (9.5.2). This consistency requirement provides a solution for the constants a, b, and f.

If one knew the constants a, b, and f, then one could use (9.5.1) and (9.5.2) to compute sample paths for k and z denoted by $\{k_{t+1}\}_{t=1}^{T}$ and $\{z_{t+1}\}_{t=1}^{T}$, given a starting condition, k_1 and z_1, and a sample path for the error terms $\{\varepsilon_{t+1}\}_{t=1}^{T}$. The sample path for the error terms, $\{\varepsilon_{t+1}\}_{t=1}^{T}$, can be drawn from a random number generator. Furthermore, if one has a sample path for k and z then it is easy to construct ones for other variables, such as output, o, or consumption, c.

Therefore, the hardest part of the problem is computing a, b, and f. These coefficients will be uncovered by linearizing the stochastic Euler equation (9.2.5). Toward this end, represent the values for k and z that would occur in a deterministic steady state by k^* and z^*. In the absence of uncertainty, the above decision rule should converge to this steady state implying

$$\ln k^* = a + b \ln k^* + f \ln z^*.$$

Thus, one can write

$$\ln k' - \ln k^* = b(\ln k - \ln k^*) + f(\ln z - \ln z^*),$$

where $\ln k - \ln k^*$ and $\ln z - \ln z^*$ denote the proportionate deviations of the capital stock and technology shock away from their deterministic steady-state values, k^* and z^*. Let

$$\widehat{k} \equiv \ln k - \ln k^* \text{ and } \widehat{z} \equiv \ln z - \ln z^*,$$

which allows the decision rule for capital to be rewritten as

$$\widehat{k}' = b\widehat{k} + f\widehat{z}. \tag{9.5.3}$$

Similarly,

$$\widehat{z}' = \rho\widehat{z} + \varepsilon'.$$

Last, the unconditional (long-run) expectations of $\widehat{k}$ and $\widehat{z}$ are given by

$$E[\widehat{k}] = E[\widehat{z}] = 0,$$

which imply that $E[\ln k] = \ln k^*$ and $E[\ln z] = \ln z^*$. To see this, note

$$\widehat{z}_{t+j} = \rho^j \widehat{z}_t + \varepsilon_{t+j} + \rho^1 \varepsilon_{t+j-1} + \cdots + \rho^{j-1}\varepsilon_{t+1},$$

so that

$$E[\widehat{z}_{t+j}|\widehat{z}_t] = \rho^j \widehat{z}_t.$$

Clearly,

$$\lim_{j\to\infty} E[\widehat{z}_{t+j}|\widehat{z}_t] = 0,$$

because $0 < \rho < 1$. Likewise,

$$\begin{aligned}\widehat{k}_{t+j} &= b^j \widehat{k}_t + f(\widehat{z}_{t+j-1} + b\widehat{z}_{t+j-2} + \cdots + b^{j-1}\widehat{z}_t) \\ &= b^j \widehat{k}_t + f(\rho^{j-1}\widehat{z}_t + b\rho^{j-2}\widehat{z}_t + \cdots + b^{j-1}\widehat{z}_t) + \varepsilon \text{ terms.}\end{aligned}$$

Thus,

$$E[\widehat{k}_{t+j}|\widehat{k}_t, \widehat{z}_t] = b^j \widehat{k}_t + f \sum_{i=1}^{j} b^{i-1}\rho^{j-i}\widehat{z}_t$$

Therefore,

$$\lim_{j\to\infty} E[\widehat{k}_{t+j}|\widehat{k}_t, \widehat{z}_t] = 0,$$

because $\lim_{j\to\infty} b^{i-1}\rho^{j-i} = 0$ for all $1 \le i \le j$, as $0 < b, \rho < 1$.

9.5.2 *Log Linearizing the Euler Equation*

Taking antilogs of (9.5.1) yields

$$z' = z^\rho \exp(\varepsilon') \tag{9.5.4}$$

Define the function $\Lambda(k, k', k'', z, \varepsilon')$ by

$$\begin{aligned}&\Lambda(k, k', k'', z, \varepsilon') \\ &\qquad = U_1(zF(k) + (1-\delta)k - k') \\ &\quad - \beta U_1(z^\rho \exp(\varepsilon')F(k') + (1-\delta)k' - k'')[z^\rho \exp(\varepsilon')F_1(k') + 1 - \delta].\end{aligned}$$

This allows the stochastic Euler equation (9.2.5) to be rewritten as

$$E[\Lambda(k, k', k'', z, \varepsilon')] = 0.$$

Note that

$$x = e^{\ln x} = \exp(\ln x), \tag{9.5.5}$$

which implies

$$\frac{dx}{d\ln x} = e^{\ln x} = x. \tag{9.5.6}$$

Hence, one can write

$$\begin{aligned}&E[\widetilde{\Lambda}(\ln k, \ln k', \ln k'', \ln z, \varepsilon')] \\ &\quad = E[\Lambda(\underbrace{\exp(\ln k)}_{k}, \underbrace{\exp(\ln k')}_{k'}, \underbrace{\exp(\ln k'')}_{k''}, \underbrace{\exp(\ln z)}_{z}, \varepsilon')] \\ &\qquad = 0.\end{aligned}$$

Furthermore, using (9.5.6) it happens that

$$\widetilde{\Lambda}_1 = \Lambda_1 k, \widetilde{\Lambda}_2 = \Lambda_2 k', \widetilde{\Lambda}_3 = \Lambda_3 k'', \widetilde{\Lambda}_4 = \Lambda_4 z, \text{ and } \widetilde{\Lambda}_5 = \Lambda_5. \quad (9.5.7)$$

This relationship will allow the subsequent analysis to compute the derivatives of Λ instead of $\widetilde{\Lambda}$.

The conjectured decision-rule (9.5.2) implies that

$$\begin{aligned}\ln k'' &= a + b\ln k' + f\ln z' \\ &= a + b\ln k' + f(\rho\ln z + \varepsilon').\end{aligned}$$

This allows the function $\widetilde{\Lambda}$ to be expressed as

$$\widetilde{\Lambda}(\ln k, \ln k', \underbrace{a + b\ln k' + f\rho\ln z + f\varepsilon'}_{\ln k''}, \ln z, \varepsilon')$$

Take a first-order Taylor expansion of the function $\widetilde{\Lambda}$ in the variables $\ln k$, $\ln k'$, $\ln z$, and ε' around the deterministic steady state where $\ln k = \ln k^*$, $\ln k' = \ln k^*$, $\ln z = \ln z^* = 0$, and $\varepsilon' = 0$. (Again, the concept of a first-order Taylor expansion is presented in Chapter A.) This is called linearizing the function $\widetilde{\Lambda}$. Let $\Lambda(*)$ denote the arguments in the function Λ are being evaluated at their values in a deterministic steady state. The above Euler equation can then be rewritten as

$$\begin{aligned}E[&\Lambda(*) + \Lambda_1(*)k^*(\ln k - \ln k^*) \\ &+ \Lambda_2(*)k^*(\ln k' - \ln k^*) + \Lambda_3(*)k^* b(\ln k' - \ln k^*) \\ &+ \Lambda_3(*)k^* f\rho(\ln z - \ln z^*) + \Lambda_3(*)k^* f(\varepsilon' - \varepsilon^*) \\ &+ \Lambda_4(*)z^*(\ln z - \ln z^*) + \Lambda_5(*)(\varepsilon' - \varepsilon^*)] = 0,\end{aligned}$$

where (9.5.7) has been used. The derivatives $\Lambda_1(*)$, $\Lambda_2(*)$, $\Lambda_3(*)$, $\Lambda_4(*)$, and $\Lambda_5(*)$ are all just constant terms. In a deterministic steady-state $\Lambda(*) = 0$, because

$$U_1(z^*F(k^*) + (1-\delta)k - k^*) - \beta U_1(z^*F(k^*) + (1-\delta)k - k^*)z^*F_1(k^*) = 0.$$

Furthermore, $E(\varepsilon' - \varepsilon^*) = 0$. Thus, the first, sixth, and eighth terms will disappear in the expression for $E[\widetilde{\Lambda}(\ln k, \ln k', \ln k'', \ln z, \varepsilon')]$. Hence,

$$\Lambda_1(*)k^*\widehat{k} + [\Lambda_2(*) + b\Lambda_3(*)]k^*\widehat{k}' + [\Lambda_3(*)k^* f\rho + \Lambda_4(*)]\widehat{z} = 0.$$

In the above equation use has also been made of the fact that $z^* = 1$. Note that expectation operator E has disappeared from the above equation, because all terms are fully known. It is as if there is no uncertainty in the economy. Thus, linearization imposes a *certainty equivalence* property. Rewrite the above equation as

$$\widehat{k}' = \underbrace{\frac{\Lambda_1(*)}{-[\Lambda_2(*) + b\Lambda_3(*)]}}_{=b}\widehat{k} + \underbrace{\frac{\Lambda_3(*)f\rho + \Lambda_4(*)/k^*}{-[\Lambda_2(*) + b\Lambda_3(*)]}}_{=f}\widehat{z}.$$

9.5.3 *Solving for the Decision Rule*

The last step is to solve for the coefficients on the conjectured decision rule or for a, b, and f. This is done in a manner similar to Chapter 6.6.9. By comparing the above equation with (9.5.3), it is obvious that

$$b = \frac{\Lambda_1(*)}{-[\Lambda_2(*) + b\Lambda_3(*)]}, \tag{9.5.8}$$

and

$$f = \frac{\Lambda_3(*) f\rho + \Lambda_4(*)/k^*}{-[\Lambda_2(*) + b\Lambda_3(*)]}. \tag{9.5.9}$$

Solving for b involves computing the solution to the quadratic equation

$$\Lambda_3(*)b^2 + \Lambda_2(*)b + \Lambda_1(*) = 0. \tag{9.5.10}$$

This equation has two roots. As will be shown below, both will be positive in value. One will be bigger than one, the other smaller. Pick the smaller one. Note that by solving (9.5.9) for f it transpires that

$$f = \frac{\Lambda_4(*)/k^*}{-[\Lambda_2(*) + b\Lambda_3(*) + \Lambda_3(*)\rho]}. \tag{9.5.11}$$

Note these expressions involve the derivatives of Λ and *not* $\widetilde{\Lambda}$. Last, a will be given by

$$a = (1-b)\ln k^* - f \ln z^* = (1-b)\ln k^*. \tag{9.5.12}$$

9.5.4 *Numerical Characterization*

To compute a numerical solution for the model, all one needs is the derivatives $\Lambda_1(*)$, $\Lambda_2(*)$, $\Lambda_3(*)$, and $\Lambda_4(*)$. These can be computed numerically, as discussed in Chapter 8. Then, the constants a, b, and f can be determined using (9.5.12), (9.5.10), and (9.5.11).

9.5.5 *Theoretical Characterization*

It will now be shown that $0 < b < 1$ and that $f > 0$.

Computing the Derivatives for $\Lambda(k, k', k'', z, \varepsilon')$

To characterize the solution theoretically, the constants $\Lambda_1(*)$, $\Lambda_2(*)$, $\Lambda_3(*)$, and $\Lambda_4(*)$ need to be calculated. By inspecting (9.2.5) and (9.5.4), it is apparent that

$$\Lambda(k, k', k'', z, \varepsilon') = U_1(\underbrace{zF(k) + (1-\delta)k - k'}_{c})$$
$$- \beta U_1(\underbrace{z^\rho \exp(\varepsilon')F(k') + (1-\delta)k' - k''}_{c'})[\underbrace{z^\rho \exp(\varepsilon')}_{z'}F_1(k') + 1 - \delta].$$

In a deterministic steady state $\beta[z^* F_1(*) + 1 - \delta] = 1$, $\varepsilon^{*\prime} = 0$, $\exp(\varepsilon^{*\prime}) = 1$, and $z^* = z^{*\rho} = 1$. Therefore,

$$\Lambda_1(*) = U_{11}(*)[z^* F_1(*) + 1 - \delta] = U_{11}(*)[F_1(*) + 1 - \delta] < 0,$$

$$\begin{aligned}\Lambda_2(*) = &-U_{11}(*) - \beta U_{11}(*)[z^{*\rho}\exp(\varepsilon^{*\prime})F_1(*) + 1 - \delta][z^{*\rho}\exp(\varepsilon^{*\prime})F_1(*) + 1 - \delta] \\ &- \beta U_1(*)z^{*\rho}\exp(\varepsilon^{*\prime})F_{11}(*) \\ = &-U_{11}(*) - U_{11}(*)[F_1(*) + 1 - \delta] - \beta U_1(*)F_{11}(*) > 0,\end{aligned}$$

$$\begin{aligned}\Lambda_3(*) &= \beta U_{11}(*)[z^{*\rho}\exp(\varepsilon^{*\prime})F_1(*) + 1 - \delta] < 0 \\ &= U_{11}(*) < 0,\end{aligned}$$

$$\begin{aligned}\Lambda_4(*) = &\, U_{11}(*)F(*) - \beta U_{11}(*)[z^{*\rho}\exp(\varepsilon^{*\prime})F_1(*) + 1 - \delta]F(*)\rho z^{*\rho-1}\exp(\varepsilon^{*\prime}) \\ &- \beta U_1(*)F_1(*)\rho z^{*\rho-1}\exp(\varepsilon^{*\prime}) \\ = &\, U_{11}(*)F(*)(1-\rho) - \beta U_1(*)F_1(*)\rho < 0,\end{aligned}$$

and

$$\begin{aligned}\Lambda_5(*) &= -\beta U_{11}(*)[z^{*\rho}\exp(\varepsilon^{*\prime})F_1(*) + 1 - \delta]z^{*\rho}\exp(\varepsilon^{*\prime})F(*) - \beta U_1(*)z^{*\rho}F_1(*)\exp(\varepsilon^{*\prime}) \\ &= -U_{11}(*)F(*) - \beta U_1(*)F_1(*).\end{aligned}$$

The Solution for b and f

Focus on the solution for b, as given by (9.5.8). It implies that

$$\begin{aligned}b = &\frac{U_{11}(*)[F_1(*) + 1 - \delta]}{U_{11}(*) + U_{11}(*)[F_1(*) + 1 - \delta] + \beta U_1(*)F_{11}(*) - bU_{11}(*)} \\ &= \frac{F_1(*) + 1 - \delta}{1 - b + F_1(*) + 1 - \delta + \beta U_1(*)F_{11}(*)/U_{11}(*)}.\end{aligned}$$

This can be expressed as a quadratic equation in b:

$$\{1 + F_1(*) + 1 - \delta + \beta U_1(*)F_{11}(*)/U_{11}(*)\}b - b^2 - F_1(*) - 1 + \delta = 0.$$

A quadratic equation has two roots. At $b = 0$ the lefthand side of the above formula is negative, because $-F_1(*) - 1 + \delta = -1/\beta < 0$. At $b = 1$ the lefthand side of the above equation is positive. The lefthand side becomes negative as b becomes large. Hence, there exits a value of $b < 1$ that solves the above equation and a value of $b > 1$ that does also. Clearly, when $b > 1$ the system would be unstable. So, throw this root away. Figure 9.5.1 portrays the situation. Observe from (9.5.11) that $f > 0$, because

$$\begin{aligned}&-[\Lambda_2(*) + b\Lambda_3(*) + \rho\Lambda_3(*)] \\ &= -(1 + 1/\beta)U_{11}(*) - \beta U_1(*)F_{11}(*) + (b + \rho)U_{11}(*) < 0\end{aligned}$$

and $\Lambda_4(*) < 0$. Last, note from (9.5.12) that $a > 0$ when $\ln k^* > 0$ and $a < 0$ when $\ln k^* < 1$. Thus, the model's local transitional dynamics around its deterministic steady state have been characterized.

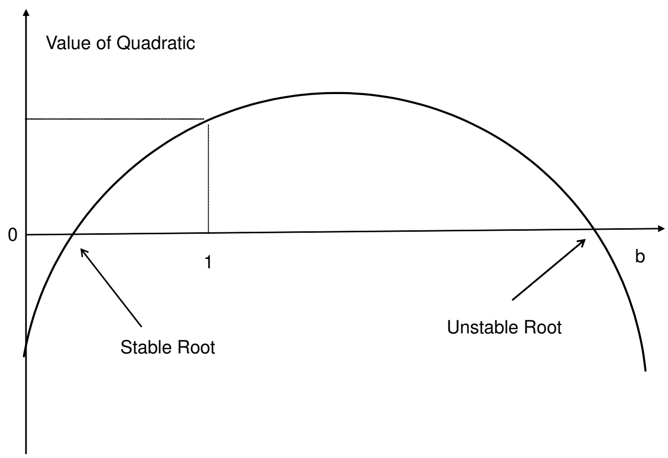


Figure 9.5.1: The two roots to the quadratic equations for b. Both roots will be positive. One will be smaller than 1, the other larger. The unstable root (the bigger one) can be discarded.

9.5.6 *Stability of the Deterministic Neoclassical Growth Model*

In the baseline version of the deterministic neoclassical growth model z is a constant. Hence, the decision for rule for capital accumulation reduces to

$$\ln k' - \ln k^* = b(\ln k - \ln k^*).$$

Additionally, around the steady state $k' - k^* \simeq k^*(\ln k' - \ln k^*)$. Therefore, one can write

$$k' - k^* = b(k - k^*).$$

Now, it has just been shown that $0 < b < 1$ so that there exists a steady state where $0 < dk_{t+1}/dk_t < 1$. The situation shown in Figure 6.7.2 therefore is apropos.

9.5.7 *Algorithm: Log-Linearized Model*

1. Compute the deterministic steady state for the model.

2. Conjecture a log-linear decision rule for capital of the form (9.5.2). Setup the Euler equation for the model and log-linearize it. This can be done by differentiating the Euler equation with respect to the logs of the variables and solving for the implied coefficients. The derivatives can be calculated either analytically or numerically–Chapter 8 covers numerical differentiation. This gives the decision rule (9.5.2). The roots for the quadratic equation for b can be computed in MATLAB using the ROOTS command. One root will lie between 0 and 1, the other will be bigger than 1. Discard the root bigger than 1.

3. Draw a sample of T random variables, $\{\varepsilon_{t+1}\}_{t=1}^{T}$–again, random number generation is discussed in Chapter 8. A sample of normally

distributed random variables can be obtained using the NORMRND command. (The older syntax is RANDN.) Make sure that the seed is fixed for the random number generator–this is also done with the RNG command, where an integer is picked for the seed.

4. Pick some initial period-0 starting value for the capital stock and technology shock, k_0 and z_0. This could be their steady-state levels so that $k_0 = k^*$ and $E[\ln z] = \ln z^* = 0$.

5. Enter period t with a level of capital, k_t, and a technology shock, z_t. Compute the capital stock for next period as follows:
$$\ln k_{t+1} = a + b \ln k_t + f \ln z_t.$$
The technology shock for next period is computed using the relationship
$$\ln z_{t+1} = \rho \ln z_t + \varepsilon_{t+1}.$$

6. Given a sample path of capital stocks and technology shocks, $\{k_t\}_{t=1}^{T+1}$ and $\{z_t\}_{t=1}^{T+1}$, data for all other variables of interest in the model, say consumption, investment and GDP, can be calculated. From these variables one can then calculate a set of business cycle statistics. When doing this take the logarithm of variable. In MATLAB standard deviations can be computed using the STD command. Likewise, correlations can be calculated using the CORRCOEF command.

9.6 *Coleman's Policy-Function Algorithm*

Another way to proceed is to solve the Euler equation (9.2.5) directly for the policy function (9.2.2). To do this, update the decision rule $k' = K(k,z)$ to get $k'' = K(k',z')$. Equation (9.2.5) can be rewritten as

Wilbur John Coleman II developed the algorithm as part of his Ph.D. thesis at the University of Chicago in 1987. It is published in Coleman (1991).

$$U_1\Big(zF(k) + (1-\delta)k - k'\Big)$$
$$= \beta \int U_1\Big(z'F(k') + (1-\delta)k' - K(k',z')\Big)[F_1(k',z') + 1 - \delta]dG(z'|z).$$

The idea for policy-function iteration is to make a guess for the decision rule K. Denote the guess to be used at stage $j+1$ by $k' = K^j(k,z)$. Then solve the equation shown below to obtain a revised guess, $k' = K^{j+1}(k,z)$.

$$U_1\Big(zF(k) + (1-\delta)k - k'\Big)$$
$$= \beta \int U_1\Big(z'F(k') + (1-\delta)k' - K^j(k',z')\Big)[F_1(k',z') + 1 - \delta]dG(z'|z).$$

Note the use of the guess $K^j(k',z')$ for k''.

To this end, let the shock process follow a m-state Markov chain where $z \in \mathcal{Z} = \{z_1, z_2, ..., z_m\}$. Denote the odds of transition from state i to state l by $\pi_{il} \equiv \Pr[z' = z_l | z' = z_i]$.

9.6.1 Algorithm: Policy-Function Iteration

Suppose that the policy function $k'' = K(k', z')$ can be approximated by some class of functions constructed over a grid $\mathcal{K} \in \{k_1, k_2, ..., k_n\}$ spanning the interval $[0, \overline{K}]$. Let $K^j(k_h, z_j)$ be a guess, within this class of functions, to be used on iteration $j+1$ for the decision rule $K(k,z)$ *at the point* $(k_h, z_j) \in \mathcal{K} \times \mathcal{Z}$. There are $n \times m$ such guesses. Unlike discrete-state-space dynamic programming, the values for the model's capital stock are not restricted to be an element of the set $\mathcal{K}$. Therefore, values for $K^j(k, z_i)$ when $k \notin \mathcal{K}$ are also required.

How are values for $K^j(k, z_i)$ determined when k is *not* a grid point? This is done by employing an interpolation scheme. Specifically, imagine that there is some interpolation scheme where a *continuous* function can be fitted, for each z_i, through the n points $\left(k_h, K^j(k_h, z_i)\right)$. With an abuse of notation, denote this interpolated function by $K^j(k, z_i)$. There are many ways to construct an interpolated function, as is discussed in Chapter 8. One could use piecewise linear interpolation, cubic spline interpolation, or interpolation with radial basis functions.

The Algorithm-Coleman (1991)

1. Enter iteration $j+1$ with a guess for $K(k, z_i)$ denoted by $K^j(k, z_i)$. The task is to compute a revised guess for $K(k, z_i)$, denoted by $K^{j+1}(k, z_i)$. To this end, at each point $(k_h, z_i) \in \mathcal{K} \times \mathcal{Z}$ a value for $K^{j+1}(k, z_i)$ can be computed by solving the equation below for k'.

$$
\begin{aligned}
&U_1\Big(F(k_h, z_i) + (1-\delta)k_h - k'\Big) \\
&= \beta \sum_{l=1}^{m} U_1\Big(F(k', z_l) + (1-\delta)k' - K^j(k', z_l)\Big)[F_1(k', z_l) + 1 - \delta]\pi_{il},
\end{aligned}
$$

   where $\pi_{i,l} = \Pr[z' = z_l | z = z_i]$. Solving for k' usually involves computing the solution to a nonlinear equation. Note that in general $k' \notin \mathcal{K}$; i.e., k' is not restricted to be a grid point. The continuity of K^j is important for solving this nonlinear equation numerically. The interpolation schemes in Chapter 8 will result in a continuous function for K^j. For the initial guess for the policy function, Coleman set investment to zero; i.e., $K^0(k, z_i) = 0$. This corresponds to assuming that Robinson Crusoe consumes all of his resources in the final period of life. The idea here is that iteration 1 corresponds to the final period of life, iteration 2 is the penultimate period of life, iteration 3 the second to last period, and iteration j to the jth last period, etc. That is, the iteration procedure can be thought of as solving the Euler equation *backwards* in time from some terminal period.

2. Compute $\rho(K^j, K^{j+1})$. If $\rho(K^j, K^{j+1}) < \varepsilon$ then stop, as convergence has been obtained. Otherwise, return to step 1 using the revised guess. The situation is shown in Figure 9.6.1 for the case of piecewise linear interpolation, which is discussed in Chapter 8.

3. Draw a sample of T uniform random variables, $\{\varepsilon_t\}_{t=1}^T$–random number generation is discussed in Chapter 8. In MATLAB a sample of uniformly distributed random variables on the $[0, 1]$ interval can be drawn using the RAND command. Make sure that the seed is fixed for the random number generator–this is done with the RNG(SEED), where SEED is some natural number. This line should be inserted just before the call for the random numbers.

4. Enter period t with a level of capital, $k_t = k_i \in \mathcal{K}$, and some past value for the technology shock, $z_{t-1} = z_r \in \mathcal{Z}$. The evolution of the technology shock, z, can be simulated using a Monte Carlo procedure–see Chapter 8, Section 8.12. In period t the technology shock may randomly transist from its past value, $z_{t-1} = z_r \in \mathcal{Z}$, to another value $z_t = z_s$ in the current period t. Now, take ε_t from the sample of random variables. Compute the current technology shock, z_t, follows:

$$z_t = z_s, \text{ if } \varepsilon \in [\sum_{u=0}^{s-1} \pi_{r,u}, \sum_{u=0}^{s} \pi_{r,u}], \text{ for } s = 1, \cdots, n,$$

where $\pi_{r0} \equiv 0$ and $\sum_{u=1}^{n} \pi_{ru} = 1$. See Figure 8.12.2 in Chapter 8 for the idea. Compute next period's capital stock, $k_{t+1} \in \mathcal{K}$, using the interpolated decision rule for capital

$$k_{t+1} = K(k_t, z_t).$$

5. For the starting value of capital just take $k_1 = k^*$, where k^* is level of capital in the deterministic steady state. For z_0 take a central value for the technology shock.

6. Given a sample path of capital stocks and technology shocks, $\{k_t\}_{t=1}^{T+1}$ and $\{z_t\}_{t=1}^{T+1}$, data for all other variables of interest in the model, say consumption, investment and GDP, can be calculated. Sometimes researchers throw away some numbers at the beginning, say $\{k_t\}_{t=1}^{n}$. This way the sample is not influenced by the starting values for k_1 and z_0. From these variables one can then calculate a set of business cycle statistics. When doing this take the logarithm of variable. In MATLAB standard deviations can be computed using the STD command. Likewise, correlations can be calculated using the CORRCOEF command.

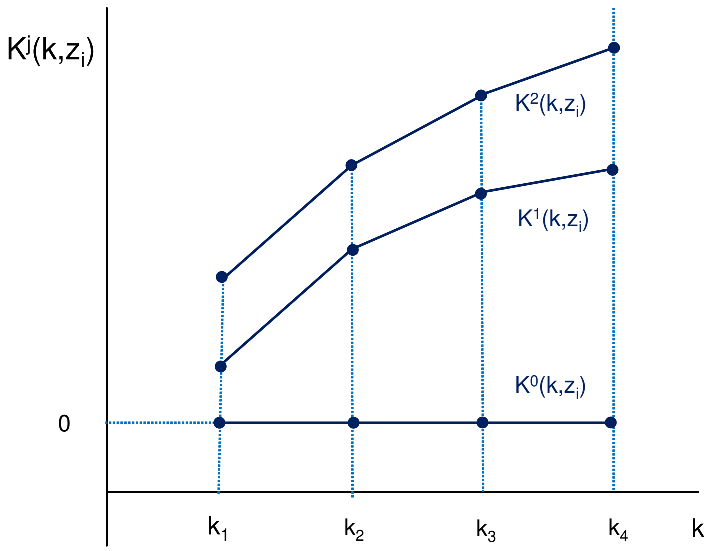


Figure 9.6.1: Coleman (1991) Algorithm with Linear Interpolation. The initial guess sets $K^0(k, z_i) = 0$, which implies Robinson Crusoe is consuming all of the resources at his disposal. This initial guess ensures that the sequence of policy functions $\{K^j(k, z_i)\}_{j=0}$ converges in a monotonic manner. Unlike discrete-state-space dynamic programming, the initial guess is important for smooth convergence when using the Coleman algorithm.

9.7 *Carroll's Endogenous Grid Method*

The Coleman method defines an exogenous grid for the current stock of capital k. This is then used to compute an interpolated decision rule $k' = K(k, z)$ by calling a nonlinear equation solver to solve the Euler equation at each grid point, given a value for the technology shock. The endogenous grid method proposed by Carroll (2006) runs the process in reverse. A fixed grid, $\mathcal{K}'$, for *next* period's capital stock k' is imposed. The values for $k' \in \mathcal{K}'$ are used to construct an interpolated consumption function, $C(x, z)$, where $x = F(k, z) + (1 - \delta)k$ is the income at the agent's disposal. The current levels of disposable income, x, that justify the choice of next period's capital stock, k', along the grid, $\mathcal{K}'$, are computed using the Euler equation, contingent on the current technology shock, z. This is the sense that Coleman's process is run in reverse. At each stage in an iteration of the algorithm, current disposable incomes and technology shocks are used to construct an interpolated consumption function, $c = C(x, z)$. This is updated to forecast next period's consumption. In sum, the current period shock, next period's capital stock, and an interpolated consumption function for next period's consumption are taken as given, and then a level of disposable income in the current period is solved for that ensures the Euler equation holds. Depending on the context, this may avoid the costly use of a nonlinear equation solver by making disposable income, x, vary endogenously. One assumption is needed to make the algorithm practicable: the marginal utility of consumption must be invertible.

9.7.1 *Algorithm: Endogenous Grid Method*

As above, define a grid for next period's capital stock $\mathcal{K}' \equiv \{k'_1, k'_2, ..., k'_n\}$ spanning the interval $[0, \overline{K}]$. Let the shock process follow a first-order m-state Markov chain, where $z \in \mathcal{Z} = \{z_1, z_2, ..., z_m\}$ and $\pi_{il} \equiv \Pr[z' = z_l | z' = z_i]$ corresponds to the probability of transitioning from state i to state l. The following algorithm uses the interpolation schemes discussed above, but in some circumstances avoids the use of a nonlinear equation solver in each iteration.

1. Enter iteration $j+1$ with a guess for next period's consumption function, $c' = C^j(x_{k_h,z_l}, z_l)$, where $x_{k_h,z_l} \equiv F(k_h, z_l) + (1-\delta)k_h$ is next period's disposable income and z_l is next period's shock. Given a value for z_l, there is a unique one-to-one relationship between x_{k_h,z_l} and k_h.

2. For each pair of next period's capital stock and current technology shock, $(k_h, z_i) \in \mathcal{K}' \times \mathcal{Z}$, a value for current disposable income x_{k,z_i}, given technology shock z_i, is recovered by solving the Euler equation

$$U_1\left(x_{k,z_i} - k\right) = \beta \sum_{l=1}^{m} U_1\left(C^j(x_{k_h,z_l}, z_l)\right)\left[F_1(k_h, z_l) + 1 - \delta\right]\pi_{il}.$$

   For a given future values of the capital stock and technology shock, k_h and z_l, the level of next period's disposable income, x_{k_h,z_l}, is known from the formula $x_{k_h,z_l} \equiv F(k_h, z_l) + (1-\delta)k_h$. The function $C^j(x_{k_h,z_l}, z_l)$ then specifies next period's consumption. Note that to solve for the current value of disposible income, x_{k,z_i}, you do not need to know the current value of the capital stock, k. But, a value for k is *implicit* from the formula $x_{k,z_i} \equiv F(k, z_i) + (1-\delta)k$. The advantage of defining the Euler equation in this fashion is that often it can be solved analytically as long as the marginal utility of consumption is invertible. Specifically,

$$x_{k,z_i} = k_h + U_1^{-1}\left(\beta \sum_{l=1}^{m} U_1\left(C^j(x_{k_h,z_l}, z_l)\right)\left[F_1(k_h, z_l) + 1 - \delta\right]\pi_{il}\right), \tag{9.7.1}$$

   where $U_1^{-1}(\cdot)$ is the inverse of the marginal utility of consumption. In certain contexts it may be possible to solve this equation quickly without the use of a nonlinear equation solver. At the end of this step, there will be $n \times m$ solutions for x, one for each value of next period's capital stock, $k' \in \mathcal{K}'$, and the current technology shock, $z \in \mathcal{Z}$. This set of solutions varies by iteration. Connected with each value of x will be a value for current consumption $c = x - k'$.

3. Now use this set of $n \times m$ solutions for c, which are functions of current disposable income, x, and the current technology shock, z, to fit a new interpolated consumption function, $c = C^{j+1}(x, z)$ that is continuous in x.

4. Compute $\rho(C^j, C^{j+1})$. If $\rho(C^j, C^{j+1}) < \varepsilon$, then stop, as convergence has been obtained. Otherwise, return to step 1 using the revised guess for the decision rule.

Example 9.1. (The invertibility of $U(c)$ with ln preferences) Let $U(c) = \ln(c)$. Then, $U_1(c) = 1/c$. Equation (9.7.1) reads as

$$x_{k,z_i} = \frac{1}{\beta \sum_{l=1}^{m} \left(1/C^j(x_{k_h,z_l}, z_l)\right) \left[F_1(k_h, z_l) + 1 - \delta\right] \pi_{il}} + k_h.$$

Example 9.2. (The invertibility of $U(c)$ with isoelastic preferences) Let $U(c) = c^{1-\rho}/(1-\rho)$. Then, $U_1(c) = c^{-\rho}$. Now equation (9.7.1) appears as

$$x_{k,z_i} = \left(\frac{1}{\beta \sum_{l=1}^{m} C^j(x_{k_h,z_l}, z_l)^{-\rho} \left[F_1(k_h, z_l) + 1 - \delta\right] \pi_{il}}\right)^{1/\rho} + k_h.$$

Remark 9.1. (Initial guess) Following the idea in Coleman (1991), a good initial guess for the consumption function would be to assume that that the person consumes all of their resources. Simply set $c'_{k_h,z_l} = F(k_h, z_l) + (1-\delta)k_h$.

Remark 9.2. (Monte Carlo simulation) Business cycle statistics can be constructed using a Monte Carlo simulation, just as in the Coleman algorithm. In a nutshell the procedure is this. One will enter an arbitrary period t with known levels of the capital stock, k_t, and technology shock, z_t. Given the current state, (k_t, z_t), compute disposable income in period t, or $x_t = F(k_t, z_t) + (1-\delta)k_t$. This gives the current level of consumption, $c_t = C(x_t, z_t)$, and next period's capital stock, $k_{t+1} = x_t - C(x_t, z_t)$. Current output and investment read $o_t = F(k_t, z_t)$ and $i_t = o_t - C(x_t, z_t)$. Then, next period's technology shock, z_{t+1}, is drawn via the Monte Carlo procedure described in the discussion of the Coleman algorithm. After this the period-$(t+1)$ state, (k_{t+1}, z_{t+1}), is known and the procedure repeats itself.

Remark 9.3. (Adding labor supply) It is easy to add labor supply into the above formulation. Let the utility function be $U\Big(c - G(l)\Big)$. Hours worked, l, can be written as a function of k and z, as discussed in Chapter 6. Define an augmented consumption function by $\widetilde{c} \equiv c - G(l) = \widetilde{C}(x, z)$ and $x_{k_h,z_l} \equiv F(k_h, z_l) + (1-\delta)k_h - G(l)$.

9.8 *Parameterized Expectations Algorithm*

Yet another method to solve dynamic stochastic models is to parameterize conditional expectation functions, typically the consumption Euler equation or the household's value function, by an ordinary polynomial function. This method was introduced by den Haan and Marcet (1990). As in the Coleman method discussed above, the goal is to find the policy function for next period capital, $k' = K(k,z)$. Now, the policy function is approximated by a flexible polynomial function given a vector of coefficients ϕ, so that

$$K(k,z) \simeq \mathcal{P}(k,z;\phi)$$

and

$$\mathcal{P}(k,z;\phi) = \sum_{i=0}^{n} \phi_i p_i(k,z),$$

where $p_i(k,z)$ is some set of basis functions such as k, z, k^2, z^2, $k\,z$, etc. Given the current capital stock and productivity level, (k_h, z_i), rewrite the Euler equation

$$U_1(c(k_h,z_i)) = \beta \sum_{l=1}^{m} U_1(c(k',z_l))[F_1(k',z_l) + 1 - \delta]\pi_{il} \tag{9.8.1}$$

as a function of k' according to

$$k' = \underbrace{\beta \sum_{l=1}^{m} \frac{U_1(c(k',z_l))}{U_1(c(k_h,z_i))}[F_1(k',z_l) + 1 - \delta]k'\pi_{il},}_{\simeq \mathcal{P}(k,z;\phi)}$$

where the righthand side of the equation is the conditional expectation to be approximated as a function of k and z. The algorithm below uses a stochastic simulation method to update the polynomial, $\mathcal{P}(k,z;\phi)$, on the righthand side. The two-stage procedure draws on Judd et al. (2011).

9.8.1 *Algorithm: Parameterized Expectations Method*

Initialize the algorithm by providing an initial guess for the vector of polynomial coefficients ϕ^1, the initial state (k_0, z_0) needed for simulations, a sequence of productivity realizations $\{z_t\}_{t=1,\ldots,T}$, where T is the simulation length. The first stage proceeds as follows:

1. Enter iteration j with coefficients ϕ^j and state (k_t, z_t), and simulate the model for T periods using

$$\begin{aligned} k'(k_t,z_t) &= \mathcal{P}(k_t,z_t;\phi^j), \\ c(k_t,z_t) &= F(k_t,z_t) - k'(k_t,z_t) + (1-\delta)\,k_t. \end{aligned}$$

2. For periods $t = 0, ..., T-1$, compute the conditional expectation in (9.8.1), where

$$\begin{aligned} k''(k', z_l) &= \mathcal{P}\left(\mathcal{P}(k_t, z_t; \phi^j), z_l; \phi^j\right), \\ c'(k', z_l) &= F(k', z_l) - k''(k', z_l) + (1-\delta)\, k'. \end{aligned}$$

3. Find $\hat{\phi}$ that minimizes the prediction error ε'

$$\varepsilon' = \beta \sum_{l=1}^{m} \frac{U_1(c(k', z_l))}{U_1(c(k_t, z_t))} [F_1(k', z_l) + 1 - \delta] k' \pi_{tl} - \mathcal{P}(k_t, z_t; \phi).$$

   Here $\hat{\phi}$ can be estimated using ordinary least-squares, least-squares using a singular value decomposition, least-absolute deviations, or principal component regressions.

4. Check for convergence of the decision rule according to

$$\frac{1}{T} \sum_{t=1}^{T} \left| \frac{k'^j - k'^{j-1}}{k'^j} \right| < \rho.$$

5. If convergence is not reached, update the vector of polynomial coefficients using fixed-point iteration with damping parameter $\gamma \in (0, 1]$, or

$$\phi^{j+1} = (1-\gamma)\phi^j + \gamma\hat{\phi},$$

   and return to step 1.

The second stage of the algorithm computes the approximation errors in the Euler equation. If the approximation is accurate, the candidate vector of polynomial coefficients $\phi^\star$ is accepted. If instead the approximation is not sufficiently accurate, the first stage can be amended by using a different approximating function $\mathcal{P}$, increasing the simulation length T, and/or choosing a different norm when estimating the coefficients $\hat{\phi}$. To compute the Euler equation errors proceed as follows:

1. Draw a new set of productivity realizations to be used as test points, $\mathcal{T} \equiv \{\tilde{z}_0, \tilde{z}_1, \cdots, \tilde{z}_T\}$. Given the converged vector of coefficients $\phi^\star$, compute $\tilde{k}'(\tilde{k}_t, \tilde{z}_t) = \mathcal{P}(\tilde{k}_t, \tilde{z}_t; \phi^\star)$ for $t = 1, 2, \cdots, T$.

2. Compute the Euler equation errors at each point $\left(\tilde{k}_\tau, \tilde{z}_\tau\right)$

$$\mathcal{E}(\tilde{k}_\tau, \tilde{z}_\tau) = \beta \sum_{l=1}^{m} \frac{U_1(c(\tilde{k}', \tilde{z}_l))}{U_1(c(\tilde{k}_\tau, \tilde{z}_\tau))} [F_1(\tilde{k}', \tilde{z}_l) + 1 - \delta] \pi_{\tau l} - 1.$$

If the mean of the errors are sufficiently small, the candidate $\phi^\star$ is accepted.

9.9 A Stochastic Dynamic Monopoly Problem

The dynamic monopolist's problem presented in Chapter 6 is reformulated to the situation where the monopolist faces a random linear demand function. In particular, demand is given by

$$p_t = \alpha_t - \frac{\beta}{2} o_t,$$

where p_t is the period-t price of the product, o_t is the monopolist's output in this period, and now α_t is a stochastic demand shifter that follows the AR1 process

$$\alpha_t = \rho\alpha_{t-1} + \varepsilon_t,$$

with

$$\varepsilon_t \sim N\big((1-\rho)\overline{\alpha}, \sigma\big).$$

Demand is decreasing in price, p_t. The monopolist produces according to the quadratic cost function

$$c_t = \frac{\gamma}{2}(o_t - \kappa o_{t-1})^2,$$

where c_t is period-t total cost and o_{t-1} is the monopolist's level of output in period $t-1$. In this random world the above cost function implies that the monopolist would like to smooth out fluctuations in his output. Under the above formulation, the long-run or unconditional expected level of the demand shifter is

$$E[\alpha] = \overline{\alpha}.$$

This can be seen by noting that

$$\alpha_t = \rho^{t-1}\alpha_1 + \varepsilon_t + \rho\varepsilon_{t-1} + \cdots + \rho^{t-2}\varepsilon_2,$$

which implies that

$$E[\alpha_t|\alpha_1] = \rho^{t-1}\alpha_1 + \frac{1-\rho^{t-1}}{1-\rho}E[\varepsilon] = \rho^{t-1}\alpha_1 + (1-\rho^t)\overline{\alpha}.$$

Clearly, as $t \to \infty$, this converges to $\overline{\alpha}$.

9.9.1 The Monopolist's Dynamic Programming Problem

The monopolist's state of the world in period t is $(o_{t-1}, \alpha_t,)$; that is, he knows the past level of his output, $o_{t-1}$and the current state of demand, α_t. The monopolist's objective is to maximize the expected present value of his profits. The mathematical transliteration of this problem is the following dynamic programming problem.

$$V(o_{-1}, \alpha) = \max_{o}\{\alpha o - \frac{\beta}{2}o^2 - \frac{\gamma}{2}(o - \kappa o_{-1})^2 + \delta E[V(o, \alpha')]\},$$

where o_{-1} is last period's output and α' is next period's state of demand. The first-order condition associated with this maximization problem is

$$\underbrace{\alpha - \beta o}_{\text{MR}} = \underbrace{\gamma(o - \kappa o_{-1}) - \delta E[V_1(o, \alpha')]}_{\text{MC}},$$

which sets marginal revenue, MR, equal to expected marginal cost, MC. By differentiating the both sides of the above dynamic programming problem, while applying the envelope theorem, it is easy to deduce that

$$V_1(o_{-1}, \alpha) = \gamma\kappa(o - \kappa o_{-1}) > 0.$$

An increase in the past level of output, o_{-1}, is beneficial to the monopolist because it reduces his current costs. By updating the equation, one obtains

$$V_1(o, \alpha') = \gamma\kappa(o' - \kappa o).$$

Using this in the first-order condition for the above dynamic programming problem gives

$$\alpha - \beta o = \gamma(o - \kappa o_{-1}) - \delta\gamma\kappa\big(E[o'|o_{-1}, \alpha] - \kappa o\big). \tag{9.9.1}$$

9.9.2 *Solving the Model via the Decision Rule Approach*

Conjecture that the monopolist's decision rule has the following linear form:

$$o = \eta + \lambda\alpha + \psi o_{-1}. \tag{9.9.2}$$

Solving the model amounts to calculating the solution for the three coefficients η, λ and ψ. The long-run (or unconditional) expected level of output is given by

$$E[o] = \frac{\eta + \lambda\overline{\alpha}}{1 - \psi}.$$

This can be seen by noting that

$$o_t = \psi^{t-1}o_1 + \eta + \psi\eta + \cdots + \psi^{t-1}\eta \\ + \lambda\alpha_t + \psi\lambda\alpha_{t-1} + \cdots + \psi^{t-1}\lambda\alpha_1,$$

so that

$$E[o_t] = \psi^{t-1}o_1 + \eta + \psi\eta + \cdots + \psi^{t-1}\eta \\ + \lambda E[\alpha_t|\alpha_1] + \psi\lambda E[\alpha_{t-1}|\alpha_1] + \cdots + \psi^{t-1}\lambda\alpha_1.$$

As t become large $E[\alpha_t|\alpha_1] = \overline{\alpha}$ and the above result obtains. Now, if $o_{-1} = o = E[o]$ and $\alpha = \overline{\alpha}$, then $E[o'] = E[o]$.

Next, $E[o]$ will be calculated using the first-order condition. This allows the constant η be determined. Toward this end, using the above results in the first-order condition for the above dynamic programming problem (9.9.1) gives

$$\bar{\alpha} - \beta E[o] = \gamma\big(E[o] - \kappa E[o]\big) - \delta\gamma\kappa\big(E[o] - \kappa E[o]\big)$$
$$\bar{\alpha} - \beta E[o] = \gamma(1-\kappa)E[o] - \delta\gamma\kappa(1-\kappa)E[o],$$

which implies

$$E[o] = \frac{1}{\beta + \gamma(1-\kappa) - \delta\gamma\kappa(1-\kappa)}\bar{\alpha} = \frac{1}{\beta + \gamma(1-\kappa)(1-\delta\kappa)}\bar{\alpha}.$$

Hence, the constant η must solve

$$\eta = (1-\psi)E[o] - \lambda\bar{\alpha}. \tag{9.9.3}$$

Making use of the conjectured decision rule in the first-order condition for the dynamic programming problem yields

$$\alpha - \beta o_t = \gamma(o_t - \kappa o_{t-1}) - \delta\gamma\kappa\big(\eta + \lambda E[\alpha'|\alpha] + \psi o_t - \kappa o_t\big).$$

Therefore,

$$(\gamma + \beta - \delta\gamma\kappa\psi + \delta\gamma\kappa^2)o_t = \alpha + \delta\gamma\kappa\eta + \delta\gamma\kappa\lambda(1-\rho)\bar{\alpha} + \delta\gamma\kappa\lambda\rho\alpha + \gamma\kappa o_{t-1}.$$

This implies that

$$\psi = \frac{\gamma\kappa}{\gamma + \beta - \delta\gamma\kappa\psi + \delta\gamma\kappa^2}, \tag{9.9.4}$$

and

$$\lambda = \frac{1 + \delta\gamma\kappa\lambda\rho}{\gamma + \beta - \delta\gamma\kappa\psi + \delta\gamma\kappa^2}. \tag{9.9.5}$$

Therefore, the solution for ψ implied by (9.9.4) solves the quadratic equation

$$-\delta\gamma\kappa\psi^2 + (\gamma + \beta + \delta\gamma\kappa^2)\psi - \gamma\kappa = 0.$$

This is the same quadratic equation as for the dynamic monopoly problem. It has two roots, a stable and unstable one. Take the stable root for ψ. From (9.9.5) the solution for λ is

$$\lambda = \frac{1}{\gamma + \beta - \delta\gamma\kappa\psi + \delta\gamma\kappa^2 - \delta\gamma\kappa\rho}.$$

9.9.3 *Decision Rule Approach–pseudo code*

Some pseudo code to solve the stochastic dynamic monopolist's problem is now presented. First, the model is solved to obtain the monopolist's decision rule. This involves finding the roots of a polynomial. Then the decision rule is simulated via a Monte Carlo. The last step is to present some output.

1. Input values for the model's parameters: $\beta = 0.5$, $\gamma = 0.5$, $\kappa = 0.9$, and $\delta = 0.96$. For the AR1 process governing α, set $\overline{\alpha} = 1$, $\sigma =$0.10, and $\rho - 0.5$.

2. Compute the steady state level of output, o^*, for the deterministic version of the model using the formula

$$o^* = \frac{1}{\beta + \gamma(1-\kappa)(1-\delta\kappa)}\overline{\alpha}.$$

3. Solve the following polynomial for ψ:

$$-\delta\gamma\kappa\psi^2 + (\gamma + \beta + \delta\gamma\kappa^2)\psi - \gamma\kappa = 0.$$

   There will be two roots to this polynomial. Take the stable root, which will be the smallest one. From this solutions for λ and η can be found. Specifically,

$$\lambda = \frac{1 + \delta\gamma\kappa\lambda\rho}{\gamma + \beta - \delta\gamma\kappa\psi + \delta\gamma\kappa^2}$$

   and

$$\eta = (1-\psi)o^* - \lambda\overline{\alpha}.$$

   The decision rule (9.9.2) for output has now been obtained. In particular, $\psi = 0.3656$, $\lambda = 0.9854$, and $\eta = 0.2664$.

4. Undertake a Monte Carlo simulation. To do this, draw a sequence of 100,000 normal random variables for the innovation ε_t. Iterate on the AR1 process for α to obtain a sequence for the demand shocks, $\{\alpha_t\}_{t=1}^{100,000}$. From this a sequence for output, $\{o_t\}_{t=1}^{100,000}$, can be obtained by using the computed decision rule. Figure 9.9.1 shows how output fluctuates randomly.

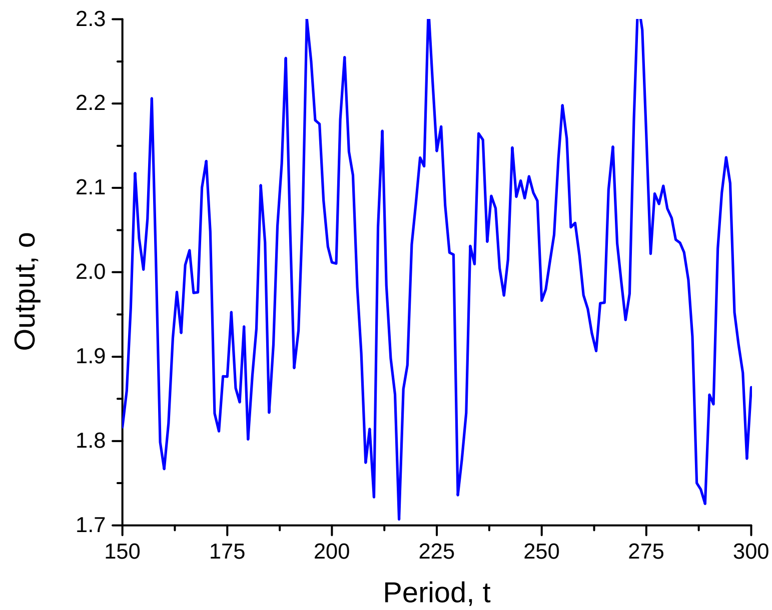


Figure 9.9.1: A representative random sample paths for output, o, that obtains from the Monte Carlo. Output fluctuations around its expected value of 2.0.

5. Compute statistics of interest, such as the standard deviations of output and prices, the correlation between output and prices, and autocorrelation for output. In the simulation $\sigma_{\ln o} = 0.0749$, $\sigma_{\ln p} = 0.1667$, $\rho_{\ln o, \ln p} = 0.8870$, and $\rho_{\ln o, \ln o_{-1}} = 0.7283$. A histogram for the stationary distribution of output is shown in Figure 9.9.2.

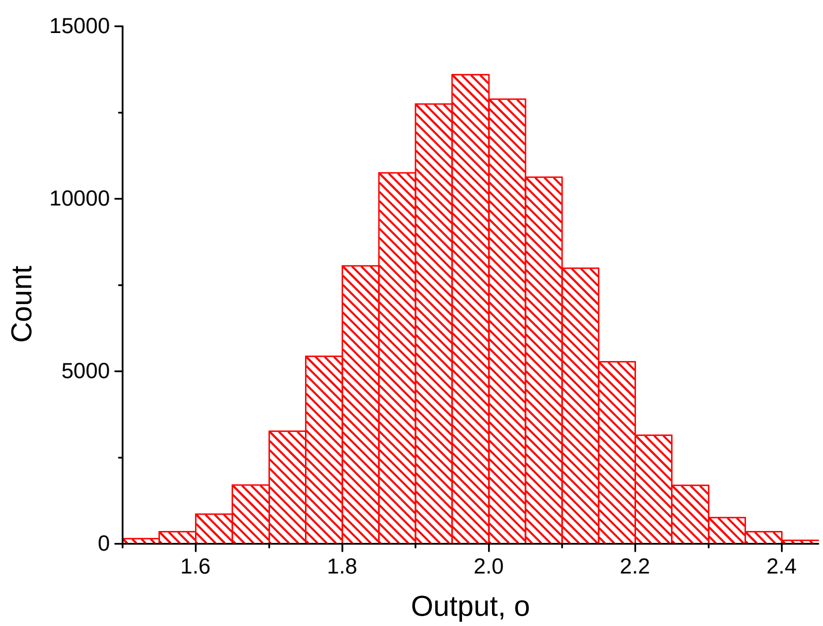


Figure 9.9.2: This is the probability distribution for output that results from the MATLAB program. It resembles a normal distribution.

9.9.4 *Solving the Stochastic Monopoly Model via Discrete-State-Space Dynamic Programming*

The setup is the same as in Section 9.9 with two changes. The stochastic demand shifter, a, is now assumed to follow a three-state Markov chain instead of an AR1 process. In particular, $\alpha \in \mathcal{A} \equiv \{\alpha_1, \alpha_2, \alpha_3\}$ with $\alpha_1 < \alpha_2 < \alpha_3$. The transition probabilities are denoted by π_{rs} for $r, s = 1, 2, 3$. Output in each period is constrained to lie on a n point grid. In particular, $o \in \mathcal{O} \equiv \{o_1, \cdots, o_n\}$.

9.9.5 *Discrete-State-Space Dynamic Programming–pseudo code*

1. Define β, γ, κ, δ as global variables. Use the same values as before.

2. Set up the three-state Markov chain for α. In particular, let $\alpha_2 = 1$, the mean value of demand shifter when a follows earlier AR1 process. Let $\alpha_1 = \alpha_2 - \Delta$ and $\alpha_3 = \alpha_2 + \Delta$. Set $\Delta = 0.0980$. Construct the transition matrix for α using Rouwenhorst's (1995) procedure to approximate an AR1 with an autocorrelation coefficient of $\rho = 0.5$, the value used before. Rousenhorst's method is discussed in Chapter 8. This gives the following symmetric transition matrix:

$$\Pi = \begin{bmatrix} 0.5625 & 0.3750 & 0.0625 \\ 0.1875 & 0.6250 & 0.1875 \\ 0.0625 & 0.3750 & 0.5625 \end{bmatrix}.$$

   Include the values for α and transition matrix, Π, in the list of global variables.

3. Build a grid of 101 equispaced points for output. Center the grid around the long-run deterministic level of output

$$o^* = \frac{\alpha_2}{\beta + \gamma(1-\kappa)(1-\delta\kappa)}.$$

   Let the grid span the interval $.9o^*$ to $1.1o^*$. Include the grid in the list of global variables.

4. Initialize a $n \times 3$ vector of guesses for the value function denoted by $V^g(o_{-1}, \alpha)$. Here o_{-1} denotes the past value of output. This vector should be included in the list of global variables.

5. Enter each iteration with a guess for the value function, $V^g(o_{-1}, \alpha)$. Compute a revised guess, $V^g(o_{-1}, \alpha)$. Inside this loop do the following.

   (a) Set up a function $M(o_{-1}, \alpha)$ that returns a n-vector of objective function values for each value of the control variable, o. The function is given by

$$M(o_{-1}, \alpha_r) = \alpha o - \frac{\beta}{2}o^2 - \frac{\gamma}{2}(o - \kappa o_{-1}) + \delta \sum_{s=1}^{3} \pi_{rs} V^g(o, a_s).$$

Here $\beta, \gamma, \kappa, \delta$ are fed in as global variables together with the demand shocks, the output grid, and the guesses for the value function.

(b) Find the value of output, o, that maximizes the function $M(o_{-1}, \alpha_r)$ for each value of lagged output, o_{-1}, and demand shock, α_r. That is, find

$$o = \arg\max M(o_{-1}, \alpha_r)$$

This gives the decision rule for output, $o = O(o_{-1}, \alpha_r)$. and a revised guess for the value function

$$V(o_{-1}, \alpha_r) = \max M(o_{-1}, \alpha_r).$$

6. Check whether $|V(o_{-1}, \alpha_r) - V^g(o_{-1}, \alpha_r)|$ is sufficiently small. If not, set $V^g(o_{-1}, \alpha_r) = V(o_{-1}, \alpha_r)$ and do another iteration. Figure 9.9.3 shows the decision rule for current output, o, for each value of three values of the shock, α, arising from computer program. Not surprisingly, the decision rules are increasing in the value of the shock, $\alpha \in \mathcal{A} \equiv \{\alpha_1, \alpha_2, \alpha_3\}$. For each shock, the decision also rises in the past value of output, o_{-1}.

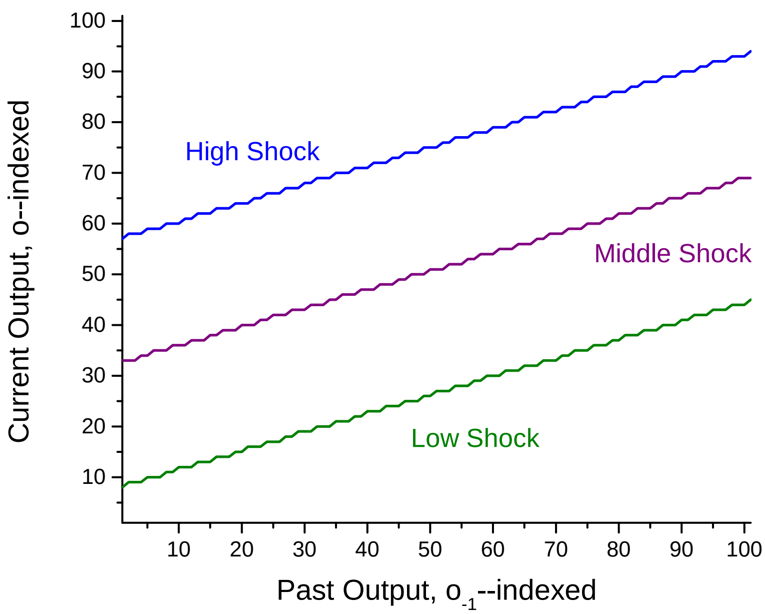


Figure 9.9.3: The decision rules for the stochastic dynamic monopolist's problem. The model is solved using discrete-state-space dynamic programming. The model was solved using a grid of 101 output levels and a 3-state Markov chain for the shock. Output is indexed by its grid point number, $1, 2, \cdots, 101$.

7. Construct the transition matrix, T, for model where

$$T = \underbrace{[p_{ir,js}]}_{3n \times 3n},$$

where

$$p_{ir,js} = Pr[o = o_j, \alpha' = \alpha_j | o_{-1} = o_j, \alpha = \alpha_i].$$

8. Compute the long-run probability vector, v^*, associated with T. This can be done by raising T to a large power–again see Chapter 8. The rows of this matrix will all be the same and correspond to the stationary distribution.

9. Compute statistics of interest using the long-run probability vector, v^*, and the transition matrix T. The stationary distribution for output, o, is displayed in Figure 9.9.4. The stationary distribution for output is derived by integrating out the demand shock to get $(v^*_{11} + v^*_{12} + v^*_{13}, \cdots, v^*_{n1} + v^*_{n2} + v^*_{n3})$. The model's stylized facts for output are very similar to those obtained using the decision rule approach. They differ a bit for prices. Specifically, one obtains $\sigma_{\ln o} = 0.04$, $\rho_{\ln p} = 0.10$, $\rho_{\ln o, \ln p} = 0.89$, and $\rho_{\ln o, \ln o_{-1}} = 0.72$. A better fit for prices could be obtained by adding more values for the demand shock.

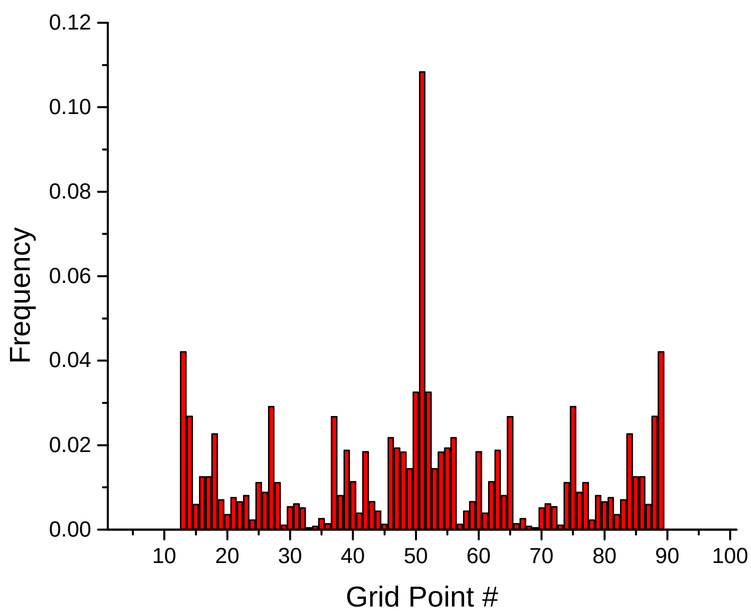


Figure 9.9.4: The stationary distribution for output associated with the transition matrix T. By adding more demand shocks this distribution could be made smoother. The extreme points on the output grid are never realized in stochastic steady state.

10. Alternatively, the stylized facts can be obtained by simulating the Markov chain via a Monte Carlo. The statistics are almost identical. Figure 9.9.5 shows the histogram for output associated with the Monte Carlo.

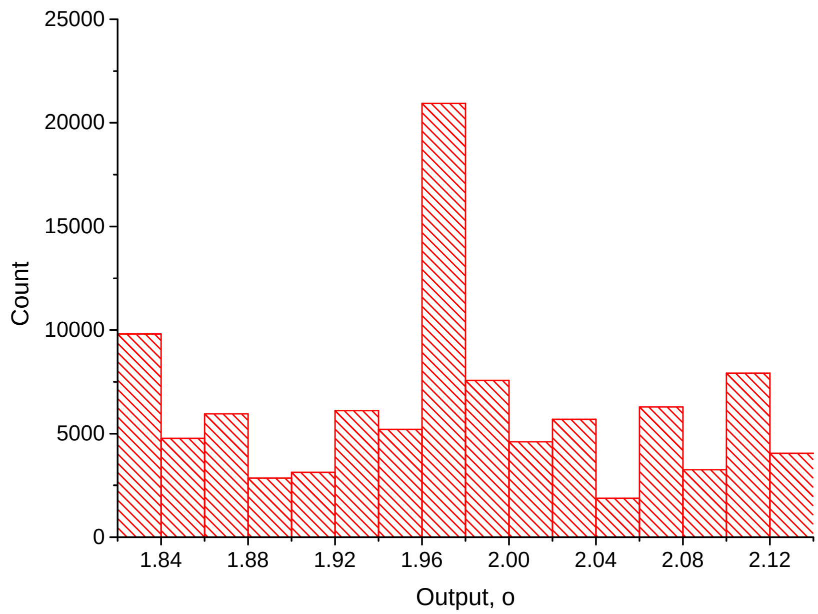


Figure 9.9.5: The histogram for output obtained by simulating the Markov chain represented by the transition matrix T using a Monte Carlo.

9.10 Problem

Consider the problem of the infinitely-lived Robinson Crusoe whose period-t utility function is given

$$U(c_t - G(h_t)), \text{ with } 0 < \beta < 1,$$

where c_t is his period-t consumption and h_t is his hours worked in period t. Let,

$$U(c - G(h)) = \frac{[c - G(h)]^{1-\gamma}}{(1-\gamma)}.$$

$$G(h) = \frac{h^{1+\theta}}{1+\theta},$$

where $\gamma, \theta > 0$. Robinson-Crusoe discounts the future at rate β where

$$\beta = \frac{1}{1+\iota},$$

and ι is his rate of time preference.

Robinson Crusoe produces output in period t, o_t, according to the following production function:

$$o_t = z_t F(k_t, h_t) = z_t k_t^{\alpha} h_t^{1-\alpha}, \text{ with } 0 < \alpha < 1.$$

where k_t is Robinson's capital stock and h_t is his work effort. The technology shock z_t follows a symmetric two-state Markov chain where

$$z_t \in \mathcal{Z} = \{z_1, z_2\},$$

with

$$z_1 = 1 - z \text{ and } z_2 = 1 + z$$

$$\Pr[z_{t+1} = z_s | z_t = z_r] = \pi_{rs},$$

and

$$\pi_{rs} = \pi_{sr}.$$

In any period t Robinson Crusoe uses some of his output for consumption and the rest for capital accumulation. Capital depreciates at rate δ over time. In any period t Robinson knows the value of the technology shock z_t.

1. Set up and solve Robinson Crusoe's dynamic programming problem. Derive Robinson's first-order condition for labor and his stochastic Euler equation governing his consumption and investment decision.

2. Suppose the economy is in a deterministic steady state (where $z_t = 1$ for all t). Derive a formula for the economy's steady-state capital stock, k^*, and labor supply, h^*.

3. Set $\alpha = 0.3$, $\beta = 1/(1.04)$, $\delta = 0.08$, $\theta = 0.6$, and $\gamma = 1.5$. What are k^* and h^*?

4. Compute a set of business cycle statistics. Solve the model using *Carroll's endogenous grid method*. Fit a polynomial for the consumption function. Report the results for both a first- and second-order polynomial. Show your consumption functions on a graph. Which consumption function fits the best? How should your initial guess for the consumption function be chosen? To compute the business cycle statistics, you will need to determine values for π_{11} and z. Why don't you need values for π_{12}, π_{21}, and π_{22}? One can think about z as determining the standard deviation of the technology shock and π_{11} its persistence–see Chapter 8. These values should picked so that the model gives the same standard deviation and serial correlation for output that is observed in the U.S. data. Suppose that the U.S. business cycle is characterized by the following set of stylized facts:

Business Cycle Statistics – U.S. Data

Variable	*Standard Deviation*	*Correlation*	*Autocorrelation*
Output	3.5	1.00	0.66
Consumption	2.2	0.74	0.72
Investment	10.5	0.68	0.25
Hours Worked	2.1	0.81	0.39
Productivity	2.2	0.82	0.77

Analyze the business cycle that comes out of this model. Also, plot the distribution for the capital stock. How did you pick the grid for the capital stock?

5. Show that it is possible for Robinson Crusoe to fix output at its steady-state level, o^*, for any value of z and k. (*Hint*: Use just the production function to obtain your result.) What are the statistical properties of the stabilized economy? What happens to Robinson's average level of utility? What is the intuition underlying your result?

10 Beyond the Representative Agent Framework

10.1 The Aiyagari model

10.1.1 Introduction

The Aiyagari (1994) model is a landmark in macroeconomics. The paper set out a basic heterogenous agent model that has become the starting point for studying incomplete markets and heterogeneity among people and firms more generally. So, it was one of the first papers to abandon the representative agent model. Other early related and notable work was done by Bewley (1977; 1980), Hugget (1993), Laitner (1992), and Imrohoroğlu (1989).

S. Rao Aiyagari (1951-1997) died at the relatively young age of 45 from a heart attack while playing tennis, one of his beloved activities. He never saw the impact that his model would have. Aiyagari was a brilliant person. Before obtaining a Ph.D. in economics he published a paper (with M.N. Mahanta) in the *Journal of Mathematical Physics* titled "On the Equivalence of the Einstein-Mayer and Einstein-Cartan Theories for Describing a Spinning Medium."

Aiyagari builds a version of the Brock and Mirman (1972) growth model that allows for a large number of individuals, subject to idiosyncratic risk, who cannot insure perfectly due to incomplete markets. People can only partially insure themselves against risk by borrowing or saving using one-period bonds with a safe return. There is a limit on how much an individual can borrow. Since risk is idiosyncratic in nature it washes out at the aggregate level, due to a law of large numbers, so that a deterministic steady-state equilibrium obtains. In the steady state of the Aiyargai model there will be a stationary cumulative distribution over individuals' assets, denoted here by $W(a_t)$ where a_t represents period-t asset holdings. Thus, $W(a_t)$ gives the fraction of people in the economy with an asset level less than equal to a_t. This is the wealth distribution for the economy. This stationary wealth distribution solves

$$W(a_{t+1}) = \int \mathbf{T}(a_{t+1}|a_t)dW(a_t)$$

where

$$\mathbf{T}(a_{t+1}|a_t) = \mathtt{Pr}[\widetilde{a}_{t+1} \leq a_{t+1}|\widetilde{a}_t = a_t].$$

The transition operator $\mathbf{T}$ gives the odds of an individual having an asset level next period less than or equal to a_{t+1}, contingent on having an asset level a_t today. In his analysis W is computed via a Monte

Carlo simulation. In particular, a single individual is subjected to random shocks for a long period of time. From this procedure a simulated wealth distribution obtains that approximates the theoretical stationary distribution. Aiyagari's analysis can be extended to include aggregate risk along the lines proposed by Boppart et al. (2018), as will be discussed. The idea underlying the Aiyagari model is illustrated in Section 10.2 using a simple stochastic, dynamic model of monopolistic competition. The demand for a firm's product will depend on the distribution of output across other firms in the economy.

Aiyagari (1994)'s analysis had two purposes.

1. *To study the impact of aggregation*. He showed how the savings decisions of many heterogeneous agents can be aggregated to obtain a deterministic steady-state wealth distribution. Unlike the neoclassical growth model, the real interest rate is not equal to the rate of time preference plus the depreciation. In particular, it is always smaller.

2. *To quantify the importance of idiosyncratic risk for savings*. Many researchers have conjectured that precautionary savings may account for a significant fraction of aggregate savings. The extent of such precautionary savings depends on how risk averse a person is and on how volatile the idiosyncratic shocks are.

The upshot of his analysis is:

1. The contribution of idiosyncratic risk to aggregate savings is modest. The aggregate savings rate increases by no more than 3 percentage points.

2. Access to asset markets is quite important in smoothing out earnings fluctuations. Asset markets allow an individual to cut their consumption variability by half and enjoy a welfare gain worth about 8% of GNP.

3. The model is consistent with certain features of the income and wealth distribution. Particularly, the distributions are positively skewed (median $<$ mean). Wealth distributions are more unequal than income distributions.

10.1.2 *The Setup*

In the Aiyagari model there is a distribution of consumer/workers of unit mass each characterized by a different level of resources that they can access. Each person seeks to maximize their expected lifetime utility as given by

$$E[\sum_{t=0}^{\infty} \beta^t U(c_t)].$$

An individual's labor period-t supply, $l_t \in [l_{\min}, l_{\max}]$, is an independently and identically distributed random variable drawn from the cumulative distribution function $L(l_t)$ with $E[l_t] = 1$. A unit of labor is paid the wage rate w. To insure against the randomness in labor income a person can borrow or lend at the interest rate r. A individual's assets in period t are denoted by a_t. This is negative when the person is in debt. The maximum level of debt that a person can incur is ϕ. People have different levels of asset holds because they experienced different histories of labor supply. An individual's period-t budget constraint reads

$$c_t + a_{t+1} = wl_t + (1+r)a_t.$$

The person also faces the borrowing constraint

$$a_{t+1} \geq -\phi.$$

Production in the economy is given by the constant-returns-to-scale production function

$$y_t = F(k_t, 1) \equiv Y(k_t),$$

where k_t is the period-t aggregate per-capita capital stock. The aggregate supply of labor is one because there is a continum of workers who each have $E[l_t] = 1$. The aggregate capital stock, k_t, evolves according to

$$k_{t+1} = (1-\delta)k_t + i_t,$$

where i_t is the period-t aggregate per-capita level of savings in the economy.

10.1.3 A Person's Choice Problem

Following Aiyagari, the model is analyzed in a transformed form. Define the variable $\widehat{a}_{t+1}$ by

$$\widehat{a}_{t+1} = a_{t+1} + \phi. \tag{10.1.1}$$

The variable $\widehat{a}_{t+1}$ represents the amount of cash that person can draw on either through savings and borrowing. Likewise, let z_{t+1} be given by

$$z_{t+1} = wl_{t+1} + (1+r)\widehat{a}_{t+1} - r\phi. \tag{10.1.2}$$

This represents the total amount of resources inclusive of labor income at the individual's disposal. With these two changes in variables the budget and borrowing constraints can be rewritten as

$$c_t + \widehat{a}_{t+1} = z_t,$$

and

$$\widehat{a}_{t+1} \geq 0.$$

An individual's dynamic programming problem can be cast as

$$V(z_t,\phi,w,r)=\max_{\widehat{a}_{t+1}\geq 0}\{U(z_t-\widehat{a}_{t+1})+\beta\int V(z_{t+1},\phi,w,r)dL(l_{t+1})\},$$

subject to (10.1.1) and (10.1.2). The Euler equation connected with the problem can have both an interior and a corner solution:

$$U_1(z_t-\widehat{a}_{t+1})=\beta(1+r)\int U_1(z_{t+1}-\widehat{a}_{t+2})dZ(z_{t+1}|z_t),\text{ if }\widehat{a}_{t+1}>0,$$

and

$$U_1(z_t-\widehat{a}_{t+1})\geq\beta(1+r)\int U_1(z_{t+1}-\widehat{a}_{t+2})dZ(z_{t+1}|z_t),\text{ if }\widehat{a}_{t+1}=0.$$

In the case of a corner solution, or when $\widehat{a}_{t+1}=0$, the individual would like to borrow more but they can't since they have hit the borrowing constraint. Thus, the marginal benefit of current borrowing, $U_1(z_t-\widehat{a}_{t+1})$, exceeds the expected future cost, $\beta(1+r)\int U_1(z_{t+1}-\widehat{a}_{t+2})dZ(z_{t+1}|z_t)$.

The above dynamic programming problem leads to a decision rule of the form

$$\widehat{a}_{t+1}=A(z_t,\phi,w,r). \tag{10.1.3}$$

The law of motion for resources then reads

$$z_{t+1}=wl_{t+1}+(1+r)\underbrace{A(z_t,\phi,w,r)}_{=\widehat{a}_{t+1}}-r\phi. \tag{10.1.4}$$

Figure 10.1.1 plots the typical shapes for these functions, assuming that the interest rate r lies below the rate of time preference, $\lambda\equiv 1/\beta-1$. This assumption is verified later. To begin with, focus on equation (10.1.3) which is shown in the lefthand panel of Figure 10.1.1. There exits some lower limit on resources, $\widehat{z}$, such the the individual will hit the borrowing constraint implying that $\widehat{a}_{t+1}=0$. This fact is proved below in Proposition 10.1. Above this point any increase in resources will be used both for consumption and either to write off debt or to save. Below this point, all resources are used for consumption and debt service. An increase in resources goes into consumption. The righthand side shows the associated function for z_{t+1} when evaluated at the two labor supply points $l_{t+1}=l_{\min}$ and $l_{t+1}=l_{\max}$. The function evaluated at other values of l_{t+1} will lie between these lines. The long-run value for z_{t+1}, or $E[z]$, is trapped between $z_{\min}$ and $z_{\max}$.

Proposition 10.1. *Assume that $\beta(1+r)<1$. Suppose that either $U_1(0)<\infty$ or $z_{\min}\equiv wl_{\min}-r\phi>0$. Then there is a $\widehat{z}>z_{\min}$ such that for all $z_t<\widehat{z}$, $c_t=z_t$ and $\widehat{a}_{t+1}=0$.*

Proof. First, note assume that $U_1(z_{\min})$ is finite. This implies that $V_1(z_{\min})=U_1(z_{\min}-\widehat{a}_{t+1})$ is finite also, a fact that will be established later. The proof now proceeds by contradiction. Suppose to the

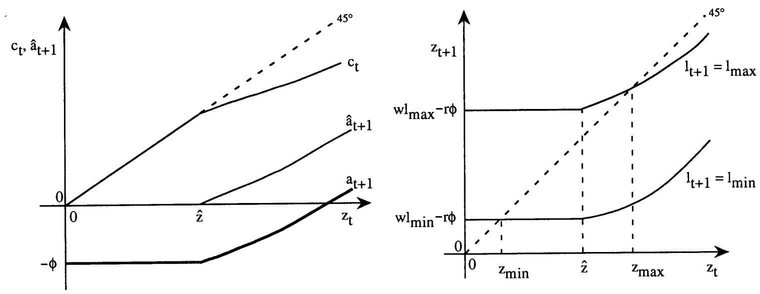


FIGURE Ia
Consumption and Assets as Functions of Total Resources

FIGURE Ib
Evolution of Total Resources

Figure 10.1.1: Figures Ia and Ib reproduced from Aiyagari (1994). The lefthand side panel shows consumption, c_t, and the cash that a person will at their disposal next period, $\widehat{a}_{t+1}$, as a function of their current resources, z_t. The borrowing constraint binds at $z_t \leq \hat{z}$. When this binds, the person will not have cash at their disposal in the future, $\widehat{a}_{t+1} = 0$, and will consume all of their current resources, $c_t = z_t$. The righthand panel shows the law of motion for resources when labor income is at its minimum and maximum values, l_{min} and l_{max}. As can be seen, a person's level of resources, z_t, will be trapped between z_{min} and z_{max}.

contrary that the borrowing constraint is not binding. Then $V_1(z_t) = U_1(z_t - \widehat{a}_{t+1}) = \underbrace{\beta(1+r)}_{<1}E[V_1(z_{t+1})] < V_1(z_{\min})$. As $z_t \to z_{\min}$ this results in a contradiction. So, there must be some neighborhood around z_m for which the proposition is true.

Second, it needs to be proved that $V_1(z_{\min})$ is finite. First, if $U_1(0)$ is finite then so will be $V_1(z_{\min}) = U_1(z_{\min} - \widehat{a}_{t+1}) \leq U_1(0)$. Second, suppose alternatively that $z_{\min} \equiv wl_{\min} - r\phi > 0$. Then, expected lifetime utility is bounded below by $U(wl_{\min} - r\phi)/(1-\beta)$. Now, $V_1(z_{\min}) = \infty$ if and only if $\widehat{a}_{t+1} = z_{\min}$. For this to be true, it must transpire that $U_1(0) \leq \beta(1+r)E[V_1(wl_{t+1} + (1+r)z_{\min} - r\phi)]$. Because, V_1 is a concave function this requires that $V_1 = \infty$ over some measurable interval, say $[z_{\min}, y]$. Now, $V(y) = V(z_{\min}) + \int_{z_{\min}}^{y} V_1(\omega)d\omega$, with $V(z_{\min}) > U(wl_{\min} - r\phi)/(1-\beta)$. If $V_1(\omega) = \infty$ over $[z_{\min}, y]$ for some y then $V(y) = \infty$, a contradiction since V is bounded from above – an assumption imposed on dynamic programming problems. □

10.1.4 *Heterogeneity and Aggregation*

Let $E[a_w]$ denote the long-run level of assets for the economy. Using (10.1.1) and (10.1.3) this is given by

$$E[a_w] = E[A(z, \phi, w, r)] - \phi.$$

Some features of this function are shown in the lefthand panel of Figure 10.1.2. Here as the rate of interest approaches the rate of time preference, $\lambda = 1/\beta - 1$, the economy's holdings of assets grow without bound. Suppose that $r = \lambda$. Then, in a world without uncertainty it would be costless for a person to hang on to assets. There is a positive probability that the individual can get a string of bad shocks. To insure against this, the person holds an infinitely large amount of assets.

In equilibrium the marginal product of capital equals its user cost. Thus,

$$Y_1(k) = r + \delta.$$

The capital-to-labor ratio is given by

$$K(r) = Y_1^{-1}(r + \delta).$$

Since the aggregate supply of labor is one this is also the per-capita demand for capital. Given the constant-returns-to-scale assumption, the wage rate, w, can be expressed as

$$w = Y(K(r)) - rK(r) \equiv W(r).$$

The economy's equilibrium is portrayed in the righthand panel of Figure 10.1.2. From the diagram it is apparent that $r < 1/\beta - 1$. This is established formally in Proposition 10.2.

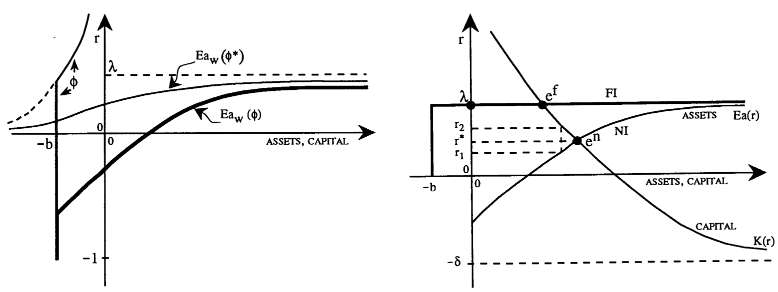


FIGURE IIa
Interest Rate versus Per Capita Assets

FIGURE IIb
Steady-State Determination

Figure 10.1.2: Figures IIa and IIb from Aiyagari (1994). The left-panel shows the aggregate level of assets, $Ea_w(\phi)$, as a function of the real interest rate, r. Aggregate assets asymptote to infinity as $r \to \lambda = 1/\beta - 1$, which is the steady-state interest rate in the deterministic growth model. The curve marked $Ea_w(\phi^*)$ gives the aggregate level of assets when the borrowing constraint is $\phi^* = wl_{min}/r$. This borrowing contraint is the maximum amount an individual could borrow but still repay with near certainty. The right-panel superimposes the demand for capital by firms, $K(r)$, on the graph. The point e^f is the equilibrium that obtains in the model with complete markets, while e^n is the equilibrium with incomplete markets. The capital stock is higher with incomplete markets. The equilibrium interest rate with incomplete markets is r^*. If $r > r^*$, then there will be an excess supply of capital, as shown by the point $r = r_2$. The reverse is true for $r = r_1$.

Proposition 10.2. *[Hugget (1997)] In a stationary equilibrium, the interest rate lies below the rate of time preference (i.e., $r < 1/\beta - 1$), provided that a measurable set of agents is borrowing constrained.*

Proof. Let $C(z)$ denote an agent's decision rule for c. In the presence of borrowing constraints

$$U_1\Big(C(z)\Big) > \beta(1+r)\int U_1\Big(C(z')\Big)dZ(z'|z). \tag{10.1.5}$$

This above equation will hold with equality for an individual who isn't borrowing constrained. The stationary distribution for z, $\mathcal{Z}(z')$ is defined by

$$\mathcal{Z}(z') = \int Z(z'|z)d\mathcal{Z}(z).$$

Now, assume that a positive mass of agents is borrowing constrained. Integrating both sides of (10.1.5) with respect to the stationary distribution gives

$$\int U_1\Big(C(z)\Big)d\mathcal{Z}(z) > \beta(1+r)\int\int U_1\Big(C(z')\Big)dZ(z'|z)d\mathcal{Z}(z)$$
$$= \beta(1+r)\int U_1\Big(C(z')\Big)d\mathcal{Z}(z').$$

This can only be true if $\beta(1+r) < 1$, or $(1+r) < 1/\beta$. □

10.1.5 *Algorithm for Computation*

Computing a solution to this model is remarkably simple. Just follow these steps:

1. Enter each iteration j with a guess for the interest rate, where $r^j < 1/\beta - 1$.

2. Compute the solution to the representative agent's dynamic programming problem assuming this guess for the interest rate.

3. Compute $E[a_w]$ by a Monte Carlo simulation of the optimal decision-rule over some large number of periods. Monte Carlo simulation is covered in Chapter 8. Aiyagari (1994) used an interpolated version of the discrete decision rule–see Chapters 8 and 9. In a stationary equilibrium, that sample path for the time series of a_t will resemble the cross-section over a_t at point in time.

4. Check excess demand in the capital market, or $K(r^j) - E[a_w]$.

   (a) Stop if $|K(r^j) - E[a_w]| <$TOLERENCE.

   (b) If excess demand is positive, then raise the interest. If it is negative, then lower it. This can be done using a bisection routine–recall Section 2.6 in Chapter 2. [In the righthand panel of Figure 10.1.2 think about $r_1 = \underline{r}$ as giving the lower bound on an iteration using the bisection routine and $r_2 = \bar{r}$ as representing the upper bound. By happenstance in the diagram, the true solution is $r^* = (r_1 + r_2)/2$.] Call the new guess r^{j+1}. Go back to step one with the new guess, r^{j+1}.

10.1.6 *Calibration*

It's time to pick functional forms for tastes, technology, and the stochastic process for labor supply. The period length is taken to be one year, so the discount factor β is set at 0.96. The model is simulated for various configurations of parameters values for these functions. Let momentary utility be represented by

$$U(c) = \frac{c^{1-\mu} - 1}{1-\mu}, \text{ where } \mu \in \{1, 3, 5\},$$

and the production technology specified as

$$Y(k) = k^{\alpha}, \text{ with } \alpha = 0.36.$$

In the analysis the model is computed for different values for the coefficient of relative risk aversion, μ. Suppose that labor income has the following first-order autoregressive representation.

$$\ln(l_t) = \rho \ln(l_{t-1}) + \sigma\sqrt{1-\rho^2}\varepsilon_t, \text{ with } \varepsilon \sim N(0,1),$$

and

$$\underbrace{\sigma}_{\text{coef var.}} \in \{0.2, 0.4\} \text{ and } \underbrace{\rho}_{\text{auto corr.}} \in \{0, 0.3, 0.6, 0.9\}.$$

Studies indicate that a reasonable value of the coefficient of variation lies between 0.2 to 0.4. The autocorrelation coefficient probably lies below 0.6. Last, it is assumed that people can't borrow which is equivalent to setting $\phi = 0$.

10.1.7 *Results: Aggregate Savings*

The impact of idiosyncratic risk on aggregate savings is moderate, at least for reasonable values of σ, ρ, and μ.

1. *Full insurance*: In the full insurance version of the model individuals are perfectly insured against shocks to their labor supply. This is the neoclassical growth model, with $r = 1/\beta - 1$. The aggregate saving rate is just $\delta k/Y(k)$. This can be rewritten as $\delta k Y_1(k)/[Y(k)Y_1(k)] = \alpha\delta/Y_1(k) = \alpha\delta/(r+\delta)$–note that capital's share of income is given by $\alpha = kY_1(k)/Y(k)$. Thus, the aggregate savings rate is not a function of μ, σ, and ρ. The full insurance baseline gives $r = 4.17$ and a savings rate of 23.67%.

2. *Moderate risk*: $\sigma = 0.4$, $\rho = 0.6$, and $\mu = 3$. Here a moderate value for the coefficient of relative risk aversion is selected. Labor shocks are not so persistent, which makes the shocks less risky because they will not persist for that long. Aggregate savings increases by 3 percentage points. Aiyagari takes this case to be at the upper end of what is reasonable.

3. *High risk*: $\sigma = 0.4$, $\rho = 0.9$, and $\mu = 5$. Now, the individual is quite risk averse. The labor shocks are quite persistent so that a bad state will endure for some time. Now there is an increase in the aggregate savings rate of about 14 percentage points.

10.1.8 *Results: Importance of Asset Trading*

The loss due to consumption variability (expressed as a fraction of consumption) is

$$-\mu\sigma_c^2/2,$$

where μ is the coefficient of relative risk aversion and σ_c is the coefficient of variation in consumption. This is easy to show. To do so, take a second-order Taylor expansion of the utility function $U(c)$ around the mean level of consumption, $\bar{c}$. This yields

$$U(c) = U(\bar{c}) + U_1(\bar{c})(c - \bar{c}) + U_{11}(\bar{c})(c - \bar{c})^2/2,$$

so that (for a small amount of risk)

$$E[U(c)] \simeq U(\bar{c}) + U_{11}(\bar{c})\bar{c}^2\sigma_c^2/2.$$

Therefore,

$$\frac{1}{\bar{c}U'(\bar{c})}\frac{dE[U(c)]}{d\sigma_c^2} = \frac{1}{2}\underbrace{\bar{c}\frac{U_{11}(\bar{c})}{U_1(\bar{c})}}_{\text{- coef. rel. aver.}} = -\frac{1}{2}\mu.$$

Consumption variability falls from 0.35 to 0.17 (for $\mu = 3, \sigma = 0.17$, and $\rho = 0.6$), when one moves from a world where agents assets are fixed at their per-capita amount to the current setting where assets may be optimally accumulated and depleted. This lead to an increase in welfare, measured in terms of consumption, of $3 \times (0.35^2 - 0.17^2)/2 \simeq 0.14$, which is about 8% of GNP.

Example 10.1. (Coefficient of Relative Risk Aversion). Let $U(c) = c^{1-\sigma}/(1-\sigma)$,with $\sigma \geq 0$. Then,

$$-c\frac{U_{11}(c)}{U_1(c)} = \sigma.$$

10.1.9 *Results: Income Distribution*

The distribution of income in the United States is unequal, as in most countries. This is reflected in the income distribution being skewed toward the right. A simple measure of this is the ratio of the median to the mean. When the distribution is skewed to the right the median level of income will be less than the mean. This transpires because the rich (or people in the upper portion of income distribution or to the right of the median person) earn a lot more than the poor, which operates to pull the mean up. So, the lower is this measure, the greater is the level of income inequality.

The income distribution can be shown by a Lorenz curve. A Lorenz curve plots the percentage of income earned by all of the population below some percentile against that percentile. Two Lorenz curves for the United States are plotted in Figure 10.1.3, one for 2009 and the other for 2019. If the distribution of income is equal, then Lorenz curve would lie on the 45° degree line. The further away the Lorenz curve is from the 45° degree line the higher is the degree of income

Max O. Lorenz (1876-1959) was an American economist. He developed the Lorenz curve in an undergraduate essay! He published a paper on this in the *Publications of the American Statistical Association* while a graduate student in 1905.

inequality. For example, in the figure the population lying below the 50th percentile accounts for far less than 50 percent of income in the United States. According to the two Lorenz curves that are plotted income inequality was worse in 2009 than in 2019.

The Gini coefficient measures the area between the Lorenz curve and the 45° degree line. The bigger this area, the more unequal is the income distribution. For a sample of incomes, $\{i_j\}_{j=1}^n$, the Gini coefficient is given by $\sum_{j=1}^n \sum_{k=1}^n |i_j - i_k| / (2n^2 \sum_{j=1}^n i_j)$. As can be seen, it measures the differences in incomes, or the i's. Normally, the Gini is thought of having a value of 0, if all incomes are equal, and a value of 1 if one person has all of the income.

The model yields positively skewed income and wealth distributions, but falls short of the amount displayed in the data.

1. *Data*:
   (a) Median income is about 80 percent of mean income.
   (b) Gini coefficient for income and wealth are 0.40 and 0.80.
2. *Model*:
   (a) Median income is over 90 percent of mean income for all parameterizations.
   (b) When $\sigma = 0.2$, $\rho = 0.6$, and $\mu = 5$, the Gini coefficients for income and wealth are 0.12 and 0.32.

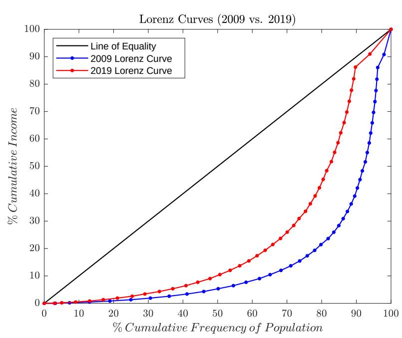


Figure 10.1.3: Lorenz curves for income in the United States, 2009 and 2019. The fact that the Lorenz curves are bowed downwards from the 45° degree line reflects inequality in U.S. incomes. This implies that the bottom x% of the population earn less than x% of aggregate income. For example, in the year 2019 the poorest 50% of population earned approximately only 10% of aggregate income.

10.2 A Stochastic, Dynamic Monopolistic Competition Problem

Consider a simple stochastic dynamic model of monopolistic competition. It builds on the dynamic model of monopoly presented in Section 6.13 and the stochastic dynamic monopoly in Section 9.9. Suppose that there is a unit mass of firms, each producing their own product. While the firms are generically the same, they each face idiosyncratic demand and survival shocks, α_t and s_t. Each firm produces according to the quadratic cost function

$$c_t = \frac{\gamma}{2}(o_t - \kappa o_{t-1})^2,$$

where c_t is period-t total cost and o_t is the monopolist's level of output in period t. Demand for a firm's product is given by

$$p_t = \alpha_t \frac{\xi}{\int_0^1 o_t(j)dj} - \frac{\beta}{2}\frac{\zeta}{\int_0^1 o_t(j)dj} o_t,$$

where p_t is the period-t price of a firm's product and $o_t(j)$ is the output of firm $j \in [0,1]$. Demand, o_t, is decreasing in the combined output of other firms, $\int_0^1 o_t(j)dj$.[1] As in Section 9.9, α_t is a stochastic demand shifter that is governed by the following AR1 process for an *incumbent* firm that survives to the next period

[1] Specifically, $o_t = 2[\alpha_t \xi - p \times \int_0^1 o_t(j)dj]/(\beta\zeta)$.

$$\alpha' = \rho\alpha + \varepsilon', \tag{10.2.1}$$

where

$$\varepsilon' \sim N\Big((1-\rho)\bar{\alpha}, \sigma\Big).$$

The realized value of this shock is specific to a firm. If a firm was guaranteed to live forever, the unconditional or long-run expectation for α is $E[\alpha] = \bar{\alpha}$–see Section 9.9.

A firm has a survival rate of s per period, which is i.i.d over time. When a firm dies it is immediately replaced by a startup. The initial value of α for a startup is $\underline{\alpha}$. A startup has zero past output as in Section 6.13. So, its output will rise over time due to a learning effect.

When choosing its own output, the firm must take into account the aggregate production by other firms. In a stationary equilibrium $\int_0^1 o_t(j)dj$ is constant, but the value of $o_t(j)$ will differ across firms due to different realizations of the demand and survival shocks. An individual firm's dynamic programming problem can be written as

$$V(o_{-1}, \alpha, s) = \max_o \{\alpha \Upsilon o - \frac{\beta}{2}\Omega o^2 - \frac{\gamma}{2}(o - \kappa o_{-1})^2 + s\delta E[V(o, \alpha', s')]\},$$

where $\Upsilon \equiv \xi / \int_0^1 o_t(j)dj$ and $\Omega \equiv \zeta / \int_0^1 o_t(j)dj$ are given by the price definition above, and δ is the firm's discount factor. The first-order condition for o is

$$\underbrace{\alpha\Upsilon - \beta\Omega o}_{\texttt{MR}} = \underbrace{\gamma(o - \kappa o_{-1}) - s\delta E[V_1(o, \alpha', s')]}_{\texttt{MC}}.$$

The lefthand side of this first-order condition is the extra revenue the firm obtains by increasing its output by a unit. The righthand side is the marginal cost of doing this. Increasing output today reduces the cost of producing output tomorrow. This future cost reduction is reflected by the term $s\delta E[V_1(o, \alpha', s')]$. This cost reduction only obtains if the firm survives into the future, which occurs with probability s. Differentiating both sides of the Bellman equation with respect to o_{-1} results in

$$V_1(o_{-1}, \alpha, s) = \gamma\kappa(o - \kappa o_{-1}).$$

Using the envelope theorem (recall s is i.i.d.), the first-order condition can now be written as

$$\alpha Y - \beta\Omega o = \gamma(o - \kappa o_{-1}) - s\delta\gamma\kappa\big(E[o'|o, \alpha', \text{firm surviving}] - \kappa o\big).$$

Conjecture a decision rule for the monopolistically competitive firm that is linear in the stochastic demand shifter and the previous period output according to

$$o = \eta + \lambda\alpha + \psi o_{-1}. \tag{10.2.2}$$

Making use of the decision rule in the first-order condition gives

$$\alpha Y - \beta\Omega o = \gamma(o - \kappa o_{-1}) - s\delta\gamma\kappa\big(\eta + \lambda E[\alpha'|\alpha, \text{firm surviving}] + \psi o - \kappa o\big).$$

This equation must hold for arbitrary values of α, o, and s. Now, $E[\alpha'|\alpha, \text{firm surviving}] = \rho\alpha + (1-\rho)\bar{\alpha}$. Rearranging terms then gives

$$\left(\gamma + \beta\Omega - s\delta\gamma\kappa\psi + s\delta\gamma\kappa^2\right) o = [s\delta\gamma\kappa\eta + s\delta\gamma\kappa\lambda(1-\rho)\overline{\alpha}] + (Y + s\delta\gamma\kappa\lambda\rho)\,\alpha + \gamma\kappa o_{-1}.$$

This implies that the terms in the conjectured decision rule are given by

$$\eta = \frac{s\delta\gamma\kappa\eta + s\delta\gamma\kappa\lambda(1-\rho)\overline{\alpha}}{\gamma + \beta\Omega - s\delta\gamma\kappa\psi + s\delta\gamma\kappa^2},$$

$$\lambda = \frac{Y + s\delta\gamma\kappa\lambda\rho}{\gamma + \beta\Omega - s\delta\gamma\kappa\psi + s\delta\gamma\kappa^2},$$

and

$$\psi = \frac{\gamma\kappa}{\gamma + \beta\Omega - s\delta\gamma\kappa\psi + s\delta\gamma\kappa^2}.$$

The solution for ψ is the stable root to the quadratic equation

$$-s\delta\gamma\kappa\psi^2 + (\gamma + \beta\Omega + s\delta\gamma\kappa^2)\psi - \gamma\kappa = 0.$$

Last, the solutions for η and λ are

$$\eta = \frac{s\delta\gamma\kappa\lambda(1-\rho)\overline{\alpha}}{\gamma + \beta\Omega - s\delta\gamma\kappa(1+\psi) + s\delta\gamma\kappa^2}$$

and

$$\lambda = \frac{Y}{\gamma + \beta\Omega - s\delta\gamma\kappa(\rho+\psi) + s\delta\gamma\kappa^2}.$$

The solution is recursive; the solution for ψ can be computed in isolation from the other parameters, to solve for λ requires knowledge of ψ, while determining η necessitates values for both ψ and λ. Furthermore, Ω and Υ are functions of the aggregate output, $\int_0^1 o_t(j)dj$. Thus, to compute the solution to a firm's problem requires information about aggregate output but computing aggregate output requires the solution to a firm's problem. This is a classic fixed point problem.

10.2.1 *Decision Rule Approach–pseudo code*

Some pseudo code to solve the stochastic dynamic monopolistic competition problem is now presented. First, make a guess for the aggregate level of output. Second, compute the firm's decision rule given this guess for aggregate output. This involves finding the roots of a second order polynomial. Simulate output using this decision rule via a Monte Carlo. Compute aggregate output. Check whether the solution for aggregate output is close to the guess. If yes, stop. If no, repeat the procedure using the computed level of aggregate output as a revised guess.

1. Set a level of tolerance for the fixed point algorithm, ε. Input values for the model's parameters: $\beta = 0.5$, $\xi = 1.4906$, $\zeta = 1.4906$, $\gamma = 0.5$, $\kappa = 0.9$, and $\delta = 0.96$. For the AR1 process governing α, set $\overline{\alpha} = 1$, $\rho_\alpha = 0.5$ and $\sigma = 0.10$. Last, set the level of α for a newborn to be $\underline{\alpha} = 0.5$ and the survival rate for an incumbant firm so that $s = 0.8$.

2. Draw a sequence of 100,000 normal random variables for the innovation ε_t and 100,000 uniform random variables for the survival shock s_t. This only needs to be done once.

3. Construct a `while` loop. Enter an iteration with a guess for aggregate output, **o**. Iterations will continue so long as this guess differs from the actual solution for output on an iteration denoted by $E[o]$. The guess for aggregate output, **o**, implies values for Υ and Ω.

4. Solve the following polynomial for ψ given **o**

$$-s\delta\gamma\kappa\psi^2 + (\gamma + \beta\Omega + s\delta\gamma\kappa^2)\psi - \gamma\kappa = 0.$$

   There will be two roots to this polynomial. Take the stable root, which will be the smallest one. With a value for ψ, solutions for λ and η can be found. Specifically, start with

$$\lambda = \frac{\Upsilon}{\gamma + \beta\Omega - s\delta\gamma\kappa(\rho + \psi) + s\delta\gamma\kappa^2}$$

   and then

$$\eta = \frac{s\delta\gamma\kappa\lambda(1-\rho)\overline{\alpha}}{\gamma + \beta\Omega - s\delta\gamma\kappa(1+\psi) + s\delta\gamma\kappa^2}.$$

   The decision rule (10.2.2) for output has now been obtained. The solutions are $\varphi = 0.3816$, $\lambda = 0.9937$, and $\eta = 0.2060$.

5. Undertake a Monte Carlo simulation using a `for` loop inside of the `while` loop. Use the shocks that have been drawn for ε and s. If $s_t \geq s$, the firm survives and α_t follows the AR1 process specified by (10.2.1). If $s_t < s$, a new firm is born and $\alpha_t = \underline{\alpha}$. Use the generated sequence for $\{\alpha_t\}_{t=1}^{100,000}$ to obtain a sequence for output, $\{o_t\}_{t=1}^{100,000}$, where $\alpha_1 = \underline{\alpha}$ and $o_0 = 0$. Compute the aggregate value of output, $E[o]$.

6. If $|E[o] - \mathbf{o}| < \varepsilon$, a solution has been found. If not, set $\mathbf{o} = E[o]$, and go back to Step 3 and update Y and Ω in the `while` loop accordingly.

7. Once the algorithm has converged, compute statistics of interest, such as the standard deviations of output and prices, the correlation between output and prices, and autocorrelation for output. In the simulation $\sigma_{\ln o} = 0.3829$, $\sigma_{\ln p} = 0.2273$, $\rho_{\ln o, \ln p} = 0.8469$, and $\rho_{\ln o, \ln o_{-1}} = 0.3759$. Figure 10.2.1 shows how output fluctuates randomly. When an incumbent firm dies it is replaced with a newborn that has a low level of initial output. This output grows over time, other things the same.

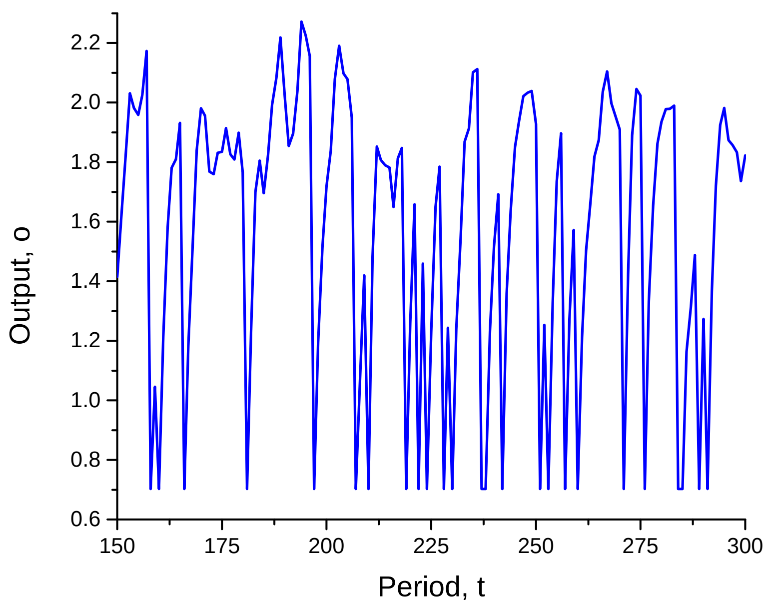


Figure 10.2.1: A representative random sample paths for output, o, that obtains from the Monte Carlo. Output fluctuates around its mean value of 1.49.

A histogram for the stationary distribution of output is shown in Figure 10.2.2. The mass point on the right is associated with the arrival of new entrants that start off with a low level of production.

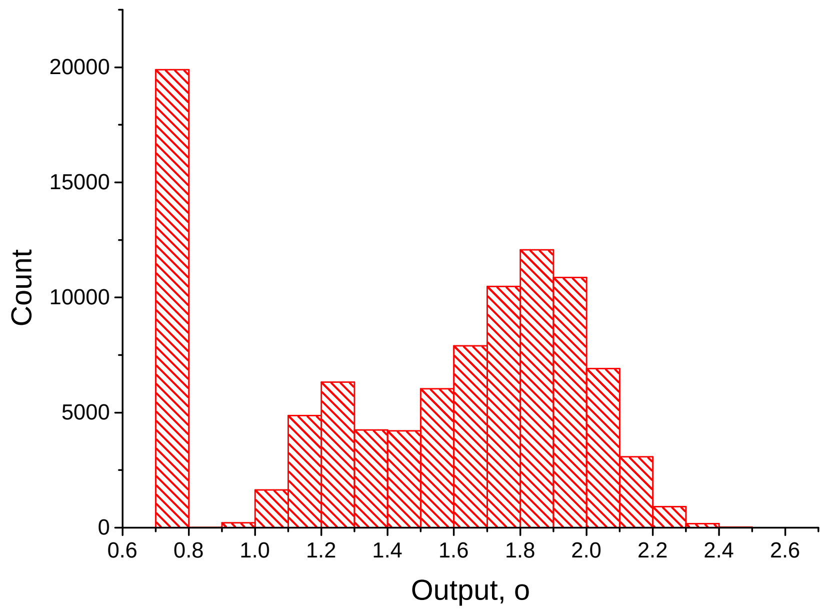


Figure 10.2.2: This is the probability distribution for output that results from the MATLAB program.

10.3 Aggregate Uncertainty

Can the model be computed with aggregate uncertainty? The answer is yes. To do so, let the aggregate production function include total factor productivity z_t so that $y_t = z_t Y(k_t)$ and assume productivity follows a first-order autoregressive process (AR(1)) in logs with ρ as the serial correlation parameter. Thus, $\ln z_t = \rho \ln z_{t-1} + \varepsilon_t$, where ε_t is white noise. The issue here is that the entire distribution of wealth is a state variable off steady state. It will change as the aggregate economy is shocked.

An algorithm for computing the Aiyagari (1994) model with uncertainty is outlined in Boppart et al. (2018). Their algorithm involves two key steps. In the first step, deterministic dynamics are computed for the model. The impulse response functions that arise from the deterministic dynamics are calculated for variables of interest, say the aggregate capital stock, aggregate consumption, the Gini coefficient, etc. These impulse response functions are only computed once. In turn, the second step involves undertaking a Monte Carlo simulation of these impulse response functions to obtain statistics of interest for the aggregate economy with uncertainty.

Timo Boppart, Per Krusell, and Kurt Mitman are economists at the Institute for International Economics in Stockholm, Sweden.

10.3.1 Deterministic Dynamics

To compute deterministic dynamics for the Aiyagari model a version of the extended path algorithm discussed in Chapter 6 can be used. Start the economy off from the steady-state wealth distribution in the

Aiyagari model. This is done using discrete-state-space dynamics programming. Here it is important to create a grid of asset holdings with more points at the lower end since the distribution will feature a large mass of individuals close to the borrowing constraint.[2] The steady-state wealth distribution is computed as the invariant distribution of the Markov chain connected with the dynamic programming problem–see Chapters 8 and 9 for a discussion of Markov chains. Then, do a one-time unforeseen shock to the innovation at the start of time. In particular, let $\varepsilon_1 = \sigma$, where σ is one standard deviation. Hence, the sequence of innovations is just $\{\varepsilon_1 = \sigma, 0, 0, 0, \cdots\}$. Thus, z will jump up or down upon impact and then return to its steady-state level (i.e. 1). In particular, $\ln z_t$ will follow the time path $\ln z_t = \rho^{t-1}\varepsilon_1$. Assume that the economy will converge back to its initial steady-state level of capital by time T.

[2] This can be achieved by defining a point in the asset grid as follows: $a_j = \underline{a} + (\overline{a} - \underline{a})\left(\frac{j-1}{n-1}\right)^{\alpha}$, where $\underline{a}$ is the minimum level of asset holdings, $\overline{a}$ is the maximum level of asset holdings, n is the total number of grid points, and α is a curvature parameter (usuallly set to 7).

1. Enter an iteration with a guess for the time path of the aggregate capital stock $\{k_t\}_{t=1}^{T}$, denoted by $\{k_t^j\}_{t=1}^{T}$. Note that a guess for the time path for the aggregate capital stock will imply a guess for the time paths for the interest and wage rates.

2. Given this guess path for aggregate capital stock, compute the representative agent's value functions and decision rules starting at period $T-1$ and then work backwards to period 1.

3. Now, use the obtained decision rule for savings and the idiosynchratic transition probabilities to simulate forward in time the evolution of the distribution of income. So, for example, in period 1 use the decision rule for saving to compute the distribution of income for period 2, and likewise in period t use the decision rule savings to compute the wealth distribution for period $t+1$. This is done using the transition matrix connected with the dynamic programming problem.

4. The wealth distribution computed for each period t implies an aggregate capital stock for that period. Thus, a sequence for the aggregate capital stock will obtain, $\{k_t\}_{t=1}^{T}$. Check $\sum_{t=1}^{T} |k_t - k_t^j| <$`tolerance`.

   (a) If so, exit the algorithm since a solution has been found.

   (b) If not, set $\{k_t^{j+1}\}_{t=1}^{T} = \{k_t\}_{t=1}^{T}$. Repeat step one using this new guess.

5. Upon convergence save the time paths (or the impulse response functions) for the variables of interest. The impulse reponse functions never need to be computed again.

10.3.2 *Simulating the Impulse Response Functions*

Let $X(j)$ represent the baseline impulse response function for the some generic variable, x, that was obtained in the previous step. The variable x is measured in logarithms as the deviation from the logarithm of its steady-state value, $x^{\star}$. That is, $100 \times X(j)$ gives the percentage deviation of the variable of interest relative to its steady-state level in the j-th period following the one-shot innovation. The impulse response function shifts up or down in proportion to the size of the unforeseen innovation. Hence, if the one-shot innovation was $\lambda\varepsilon_1$ instead of $\varepsilon_1 > 0$, with $\lambda > 1$, then the impulse response function shifts up by the factor λ.

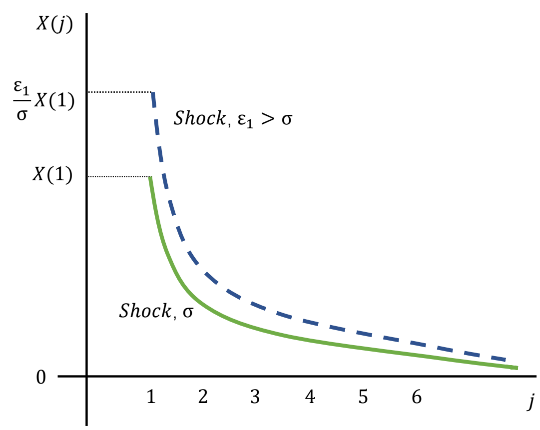


Figure 10.3.1: Impulse Response Functions: The solid green line shows the baseline response of x to an unexpected positive shock in period 1 to ε of size σ. The response is measured in terms of the gap between the logged variable and the log of its steady-state value. The dashed blue line shows what would happen if instead a larger shock, $\varepsilon_1 > \sigma$, occurred. The baseline impulse response just needs to be scaled up by the factor ε_1/σ.

The response of x_t to a sequence of innovations $\{\varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, \cdots\}$ is given by a scaled moving average of the past shocks multiplied by the baseline impulse response function obtained from the deterministic simulation, or

$$x_t = \sum_{j=1}^{J} \left(\frac{\varepsilon_{t+1-j}}{\sigma}\right) X(j).$$

In the above formula, the baseline impulse response function $X(j)$ gives the impact on x_t of an innovation of size σ that happened j periods ago. If the innovation was instead of size ε_{t+1-j}, then things need to be rescaled by $\varepsilon_{t+1-j}/\sigma$.

A Monte Carlo simulation is then used to generate the time-series process for x_t. Specifically, draw a random sequence for the ε_t's, $\{\varepsilon_t\}_{t=1}^{N}$where N is a large number. Next, simulate the changes in x_t's as follows

$$\begin{aligned} x_J &= \left(\frac{\varepsilon_J}{\sigma}\right) X(1) + \left(\frac{\varepsilon_{J-1}}{\sigma}\right) X(2) + \cdots + \left(\frac{\varepsilon_1}{\sigma}\right) X(J), \\ x_{J+1} &= \left(\frac{\varepsilon_{J+1}}{\sigma}\right) X(1) + \left(\frac{\varepsilon_J}{\sigma}\right) X(2) + \cdots + \left(\frac{\varepsilon_2}{\sigma}\right) X(J), \\ \vdots &= \vdots \\ x_t &= \left(\frac{\varepsilon_t}{\sigma}\right) X(1) + \left(\frac{\varepsilon_{t-1}}{\sigma}\right) X(2) + \cdots + \left(\frac{\varepsilon_{t+1-J}}{\sigma}\right) X(J). \end{aligned}$$

Finally, compute the desired descriptive statistics from the obtained (logged) deviations from the steady state.

11 Neural Networks for Economic Problems

11.1 Neural Networks

Neural networks came to prominence in the 1940s to study how neurons interacted in a human brain. Much progress has been made since then. Neural networks are currently being used to recognize patterns in various contexts (text, speech, images, data), which are then classified into common structures that can be used for prediction. One of the recent applications of these tools is in solving macroeconomic models. A neural network consists of three types of layers: (i) an input layer, which takes as input the states of the economic problem (e.g., today's capital and productivity, k_t and z_t) and might also take parameters to be estimated; (ii) hidden layers, which perform nonlinear transformations of the input layer or previous hidden layers if these exist; and (iii) an output layer, which summarizes the policy functions of interest (e.g., tomorrow's capital, k_{t+1}) as well as the estimated parameters.

Each layer is comprised of several nodes or neurons. Layers are then connected to each other through a combination of weights applied to the output of the nodes in other layers, which are used as inputs. There are multiple connection patterns. For instance, layers are fully connected if every node in a layer is connected to every node in the other layer. There are two types of neural networks: (i) the feedfoward neural network, which is a network that only proceeds forward across layers; and (ii) the recurrent neural network, which proceeds both forward and backward across layers and hence allows for feedback loops. Figure 11.1.1 below depicts a fully connected feedfoward neural network, with three nodes in the input layer, two hidden layers with four nodes each, and an output layer with two nodes.

How can neural networks be used in macroeconomics? Neural networks can be used to approximate the equilibrium functions of an economic model (e.g., Euler equations, labor supply decisions, market clearing conditions) and derive policy and value functions. The implementation of a neural network to an economic model involves four steps: (1) defining a class of function approximators, (2) defining a loss function that measures the quality of the approximation, (3) defining an updating mechanism to improve the approximation, and (4) defining a sampling method to update the states.

1. Function approximator

   The function approximator that we choose is a fully connected feed-forward neural network. Let $\mathcal{N}_{\boldsymbol{\rho}}$ define a neural network where $\boldsymbol{\rho}$ is a vector of trainable parameters and $\boldsymbol{x} := \left[\boldsymbol{z}^T, \boldsymbol{k}^T\right]^T$ is a matrix of states that includes a vector of productivity shocks $\boldsymbol{z} \in \mathcal{Z}$ and of capital $\boldsymbol{k} \in \mathbb{R}$ (here T stands for the transpose). The neural network satisfies the following mapping

   $$\boldsymbol{x} \to \mathcal{N}_{\boldsymbol{\rho}}(\boldsymbol{x}).$$

   Through its different layers, the neural network applies nonlinear transformations to the vector of states. Let the network have K layers and m_i nodes within each layer $i = \{1, ..., K\}$. The network's vector of trainable parameters $\boldsymbol{\rho}$ is a high-dimensional object that includes matrices of weights $\{\boldsymbol{W}_1, ..., \boldsymbol{W}_K\}$ and bias vectors $\{\boldsymbol{b}_1, ..., \boldsymbol{b}_K\}$ for the K layers of the network.

   The output layer of the neural network is such that

   $$\mathcal{N}_{\boldsymbol{\rho}}(\boldsymbol{x}) = \sigma_K\left(\boldsymbol{W}_K ... \sigma_2\left(\boldsymbol{W}_2 \sigma_1\left(\boldsymbol{W}_1 \boldsymbol{x} + \boldsymbol{b}_1\right) + \boldsymbol{b}_2\right) ... + \boldsymbol{b}_K\right).$$

   Here, σ_i are activation functions (e.g., sigmoid, hyperbolic tangent, rectified linear unit or ReLU) applied elementwise to each entry of a vector, i.e., $\sigma_i(\boldsymbol{x}) = [\sigma_i(x_1), ..., \sigma_i(x_{m_i})]^T$, $\boldsymbol{W}_i \in \mathbb{R}^{m_i \times m_i}$ and $\boldsymbol{b}_i \in \mathbb{R}^{m_i}$ are the weight matrix and bias vector of layer i across the different m_i nodes.

2. Loss function

   The loss function measures the quality of the approximation to the economy's equilibrium, by penalizing deviations from the equilibrium conditions. It can be written as the mean squared error of all equilibrium conditions according to

   $$Loss(\boldsymbol{\rho}, \mathcal{D}) = \frac{1}{|\mathcal{D}|} \sum_{\boldsymbol{x}_j \in \mathcal{D}} \left[\left(\epsilon_{\boldsymbol{x}_j}^{\text{Euler}}(\boldsymbol{\rho})\right)^2 + ... + \left(\epsilon_{\boldsymbol{x}_j}^{\text{Market Clearing}}(\boldsymbol{\rho})\right)^2\right],$$

   where $\mathcal{D}$ refers to the training dataset (more on this below), which contains the set of states $\boldsymbol{x}$ at which the equilibrium conditions are evaluated, ϵ^{Euler} is a vector containing the Euler equation errors when evaluated at $\boldsymbol{x}$, and $\epsilon^{\text{Market Clearing}}$ is the errors from market clearing conditions when evaluated at $\boldsymbol{x}$.

3. Updating mechanism

   The goal is to find the set of trainable parameters of the neural network, $\boldsymbol{\rho}$, such that the resulting approximation of the policy functions satisfies the equilibrium conditions. A rule to update $\boldsymbol{\rho}$ is thus needed. These parameters are updated following a gradient descent algorithm that goes in the direction in which the loss function declines. For each element $j \in \{1, ..., \text{length}(\boldsymbol{\rho})\}$, the new guess of the trainable parameters satisfy the following

$$\rho_j^{\text{new}} = \rho_j^{\text{old}} - \alpha^{\text{learn}} \frac{\partial Loss(\boldsymbol{\rho}^{\text{old}}, \mathcal{D})}{\partial \rho_j^{\text{old}}},$$

   where $\alpha^{\text{learn}} > 0$ is the learning rate or step size in each gradient descent step. Note that the gradient of the loss function is computed using automatic differentiation for each element of the network's learning parameters in the previous iteration $\boldsymbol{\rho}^{\text{old}}$.

4. Sampling and training data

   The neural network relies on a large amount of "data." The more data there is to train the network, the better, and faster, will the approximation be. The idea here is to use the economic model as a data generating process. Start with a randomly set of neural network parameters $\boldsymbol{\rho}$ and an arbitrary vector of states $\boldsymbol{x}_1^0$ (here 1 corresponds to the first observation of the vector of states and 0 is the first iteration of the training dataset). Then simulate the economy for $T-1$ periods based on the approximated equilibrium functions derived from the neural network. The resulting T simulated states correspond to the first training dataset $\mathcal{D}^0 = \{\boldsymbol{x}_1^0, ..., \boldsymbol{x}_T^0\}$, which can be used to solve for the equilibrium conditions and to update the neural network parameters as in step 3.
   In the next iteration, the states used in the previous iteration serve as the initial guess, i.e., $\boldsymbol{x}_1^1 = \boldsymbol{x}_T^0$. Using the updated neural network parameters, $\boldsymbol{\rho}^{\text{new}}$, a new sample path of size $T-1$ is obtained consisting of the updated training dataset $\mathcal{D}^1 = \{\boldsymbol{x}_1^1, ..., \boldsymbol{x}_T^1\}$. With this training dataset, the equilibrium conditions are evaluated, the loss function is computed, and the neural network parameters are updated.

11.1.1 *Algorithm*

1. Define the network structure, the number of input, hidden, and output layers, the number of nodes within each layer as well as the activation functions to be applied in each layer. Choose also the learning rate for the updating of the neural network parameters,

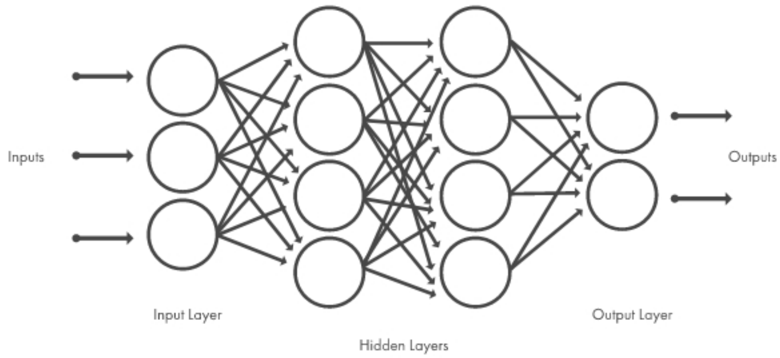


Figure 11.1.1: A simple feedforward neural network

α^{learn} and the length for the simulation. In this step, the neural network parameters, $\boldsymbol{\rho}$, are initialized.

2. Generate a training dataset for the states, by computing vectors $\{z_1^0, ..., z_T^0\}$ and $\{k_1^0, ..., k_T^0\}$.

3. Enter simulation s with training dataset $\mathcal{D}^{s-1}$ and the neural network parameters $\boldsymbol{\rho}^{s-1}$.

   (a) Generate a new training dataset, $\mathcal{D}^s$, by simulating z with new productivity shocks and k' with the capital's law of motion.

   (b) Use the updated training dataset to solve the model's equilibrium conditions and retrieve the loss function $Loss(\boldsymbol{\rho}^{s-1}, \mathcal{D}^s)$.

   (c) Use automatic differentiation to compute the gradient of the loss function for every element of the network's trainable parameters, $\frac{\partial Loss(\boldsymbol{\rho}^{s-1}, \mathcal{D}^s)}{\partial \rho_j^{s-1}}$ and get the updated parameters $\boldsymbol{\rho}^s$.

4. Use the converged neural network parameters, $\boldsymbol{\rho}^N$, to retrieve the policy and value functions, and compute statistics of interest.

11.2 *Activation functions*

The commonly used activation functions are

- Rectified Linear Units (ReLU) functions feature a kink and are given by $f(x) = \max\{0, x\}$.

- Leaky ReLU functions are given by $f(x) = \begin{cases} x & \text{if } x \geq 0 \\ \alpha x & \text{if } x < 0 \end{cases}$, where $\alpha < 1$.

- Scaled Exponential Linear Units (SELU) functions are given by $f(x) = \begin{cases} \lambda x & \text{if } x \geq 0 \\ \lambda\alpha(e^x - 1) & \text{if } x < 0 \end{cases}$, where $\lambda \approx 1.0507$ and $\alpha \approx 1.6733$.

- Softplus functions are given by $f(x) = \log(1 + exp(x))$.
- Sigmoid functions are useful for variables spanning the unit interval and are given by $f(x) = \frac{1}{1+exp(-x)}$.
- Swish functions are given by $f(x) = x\text{sigmoid}(\beta x)$, where β is a learnable parameter and sigmoid() is the sigmoid activation function.
- Tanh functions are given by $f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$.
- Gaussian Error Linear Units (GELU) functions are given by $f(x) = x\Phi(x)$, where $\Phi(x)$ corresponds to the standard Normal cumulative distribution function.

11.3 *A Heterogeneous Firms Model Solved With Neural Networks*

Consider a version of the stochastic dynamic monopolistic competition model described in Section 10.2. Individuals have preferences over the consumption of different products j given by

$$\int_0^{N_t} \alpha_t(j) \frac{q_t(j)^{1-\zeta} - 1}{1-\zeta} dj,$$

where $\alpha_t(j)$ is a preference shock for the variety produced by a firm j and is governed by the following AR1 process

$$\alpha_{t+1} = \rho_\alpha \alpha_t + \varepsilon_{\alpha_{t+1}} \qquad \text{for each } j. \tag{11.3.1}$$

Here, the preference shock disturbances are normally distributed according to $\varepsilon_{\alpha_{t+1}} \sim N\Big((1-\rho_\alpha)\bar{\alpha}, \sigma_\alpha\Big)$ such that the unconditional mean of α is $\bar{\alpha}$.

Individuals face the following budget constraint

$$\int_0^{N_t} p_t(j) q_t(j) dj = y_t,$$

where y_t is income, which is subject to aggregate uncertainty and evolves according the following AR1 process

$$y_{t+1} = \rho_y y_t + \varepsilon_{y_{t+1}}. \tag{11.3.2}$$

The aggregate income shock is given by $\varepsilon_{y_{t+1}} \sim N\left((1-\rho_y), \sigma_y\right)$, where the unconditional mean of income is assumed to be equal to one.

A firm j then faces the following inverse demand function for its product

$$p_t = \frac{\alpha_t y_t o_t^{-\zeta}}{\Lambda_t},$$

where

$$\Lambda_t \equiv \int_0^{N_t} \alpha_t(j) o_t(j)^{1-\zeta} dj \tag{11.3.3}$$

is a measure related to aggregate output. Assume that a firm faces a quadratic cost function given by

$$\frac{\gamma}{2}(o_t - \kappa o_{t-1})^2.$$

A firm's dynamic programming problem can now be written as

$$V(o_{-1}, \alpha, \Lambda, y) \quad = \quad \max_o \{\tfrac{\alpha y}{\Lambda} o^{1-\zeta} - \tfrac{\gamma}{2}(o - \kappa o_{-1})^2 + \delta E[V(o, \alpha', \Lambda', y')]\}$$

subject to equations (11.3.1), (11.3.2), (11.3.3), and the aggregate law of motion $\Lambda' = G(\Lambda, y)$. The firm's first-order condition is given by

$$(1-\zeta)\frac{\alpha y}{\Lambda} o^{-\zeta} = \gamma(o - \kappa o_{-1}) - \delta\gamma\kappa \left(E[o'|o, \alpha', \Lambda', y'] - \kappa o \right). \tag{11.3.4}$$

Instead of conjecturing a decision rule for the firm's output as in Section 10.2, let the law of motion of a firm's output,

$$o = g(o_{-1}, \alpha, \Lambda, y),$$

be approximated by a neural network. The neural network is also used to approximate the law of motion of the aggregate endogenous state, $\Lambda' = G(\Lambda, y)$. Now write the Euler equation as

$$\begin{aligned} (1-\zeta)\frac{\alpha y}{\Lambda} g(o_{-1}, \alpha, \Lambda, y)^{-\zeta} \quad = \quad & \gamma(g(o_{-1}, \alpha, \Lambda, y) - \kappa o_{-1}) \\ & -\delta\gamma\kappa \sum_{\alpha', y'} \Gamma(\alpha'|\alpha)\Gamma(y'|y)\, g(g(o_{-1}, \alpha, \Lambda, y), \alpha', G(\Lambda, y), y') \\ & +\delta\gamma\kappa^2 g(o_{-1}, \alpha, \Lambda, y) \end{aligned} \tag{11.3.5}$$

where $\Gamma(\alpha'|\alpha)$ and $\Gamma(y'|y)$ are the transition probabilities for the idiosyncratic preference shock α and the aggregate demand shock y.

11.3.1 *Neural networks–pseudo code*

As in Section 11.1, proceed by first defining the network's structure and the function approximator to compute the model's solution. Here, use a fully connected feedforward neural network with $(2n_f + 1)$ input layers, where n_f is the number of simulated firms for the pair of states (o_{-1}, α), and 1 for the aggregate demand shock y. Here, we abstract from using Λ in the input layer because we recover it using the network's prediction of firms' current output according to equation (11.3.3). Add two hidden layers with $200 \times (2n_f + 1)$ nodes each and n_f output layers for the current output of a firm. The activation functions between the input layers and the hidden layers are assumed to be ReLU, and softplus are the activation functions for the output layers.

Next, define the loss function. The model's solution implies that the firms' first-order conditions hold, so set the mean squared error of equation (11.3.5) as the loss function, or

$$Loss(\boldsymbol{\rho}, \mathcal{D}) = \frac{1}{|\mathcal{D}|} \sum_{\mathcal{D}} \left[LHS(\boldsymbol{\rho}, \mathcal{D}) - RHS(\boldsymbol{\rho}, \mathcal{D}) \right]^2 ,$$

where $\boldsymbol{\rho}$ are the neural network parameters, $\mathcal{D}$ are the simulated data, and LHS and RHS are the left and right-hand-side of equation (11.3.5).

The neural network parameters $\boldsymbol{\rho}$ are then updated using a gradient descent algorithm as defined above. The starting data (o_{-1}, α, y) is obtained by simulating values of idiosyncratic preference shock α and of the aggregate demand shock y. The starting values for the previous period output, o_{-1}, are recovered from the steady state Euler equation. The data is then updated using simulated draws for the disturbances for α and y, and the neural network for current period output.

The pseudo code to solve the stochastic dynamic monopolistic competition problem using neural networks is now discussed in greater detail.

First, make a guess for the aggregate level of output. Second, compute the firm's decision rule given this guess for aggregate output. This involves finding the roots of a second order polynomial. Simulate output using this decision rule via a Monte Carlo. Compute aggregate output. Check whether the solution for aggregate output is close to the guess. If yes, stop. If no, repeat the procedure using the computed level of aggregate output as a revised guess.

1. Input values for the model's parameters: $\zeta = 0.2857$, $\gamma = 0.5$, $\kappa = 0.9$, and $\delta = 0.96$. For the AR1 process governing y, set $\rho_y = 0.9$ and $\sigma_y = 0.04$. For the AR1 process governing α, set $\overline{\alpha} = 1$, $\rho_\alpha = 0.5$ and $\sigma_\alpha = 0.10$. Discretize these stochastic processes (e.g. using Gauss-Hermite quadrature) to get the nodes and weights for the aggregate and idiosyncratic shocks needed to compute the expected value of a firm's output.

2. Set the number of simulated firms $n_f = 10,000$, the number of data points $n_d = 1,000$, and the number of simulated draws $n_t = 10,000$. Draw a sequence of $(n_d \times n_t)$ normal random variables for the aggregate demand shocks, ε_{y_t}, and $(n_d \times n_f \times n_t)$ normal random vari-

ables for the idiosyncratic shocks, ε_{α_t}. This only needs to be done once for the Monte Carlo simulation. These innovations will then be used to simulate the path of α and y to train the network.

3. To initialize the neural network, generate n_d random data points for the aggregate shock y and $(n_d \times n_f)$ points for the idiosyncratic shock α. To generate $(n_d \times n_f)$ random points of the previous period output, o_{-1}, use the steady state Euler equation evaluated at the randomly generated values of α and y, i.e., $o^{\star} = \left[\frac{(1-\zeta)\alpha y}{\alpha\gamma(1-\kappa)(1-\delta\kappa)}\right]^{1/2}$, where it was assumed that $\Lambda = \alpha o^{1-\zeta}$.

4. Define the neural network architecture: $(2n_f + 1)$ hidden layers for the states (o_{-1}, α, y); 2 hidden layers with $20 \times n_f$ nodes each; and 1 output layer for firms' output of size n_f. Declare that layers are fully connected and specify the ReLU as the activation functions to be applied across each layer. For the output layer use a softplus activation function to avoid getting output at zero. Define the number of episodes (or simulations) to assess the convergence of the neural network, say 20,000. Set the learning rate of the network's trainable parameters to 0.0003.

5. Construct a `for` loop over the number of episodes. Enter iteration j by generating training data for the network: take the inputs from the previous iteration $(o_{-1}^{j-1}, \alpha^{j-1}, y^{j-1})$ as starting values; then for every $t = \{1, ..., n_t\}$, compute α_{t+1} and y_{t+1} using α^{j-1}, y^{j-1}, and the draws of ε_{α_t} and ε_{y_t}, and use the neural network to predict current output o_t. Note that at each $t$the neural network is updated with the simulated inputs $(o_t^{j-1}, \alpha_{t+1}^{j-1}, y_{t+1}^{j-1})$.

6. Armed with new guesses of inputs $(o_{-1}^{j}, \alpha^{j}, y^{j})$ of size $((2n_f + 1) \times n_d \times n_t)$, solve the model for each point. The expected output is computed using the neural network and the discretized shock processes. Compute the Euler equation (11.3.5) for each $(o_{-1}^{j}, \alpha^{j}, y^{j})$. The loss function to assess the quality of the approximation is the mean squared error in the Euler equation evaluated at $(2n_f \times n_d \times n_t)$ points.

7. Use automatic differentiation to compute the gradient of the loss function with respect to the neural network's trainable parameters.

8. Update the neural network's trainable parameters using a gradient-based algorithm (e.g. the adaptive moment estimation (Adam) algorithm).

9. Once the algorithm has converged, compute statistics of interest, such as the standard deviations of output and prices, the correlation between output and prices, and autocorrelation for output.

Figure 11.3.1 depicts the mean squared error associated with the Euler equation (11.3.5) across simulations. The neural network is more accurate with more simulations. Figure 11.3.2 displays the probability distribution of output.

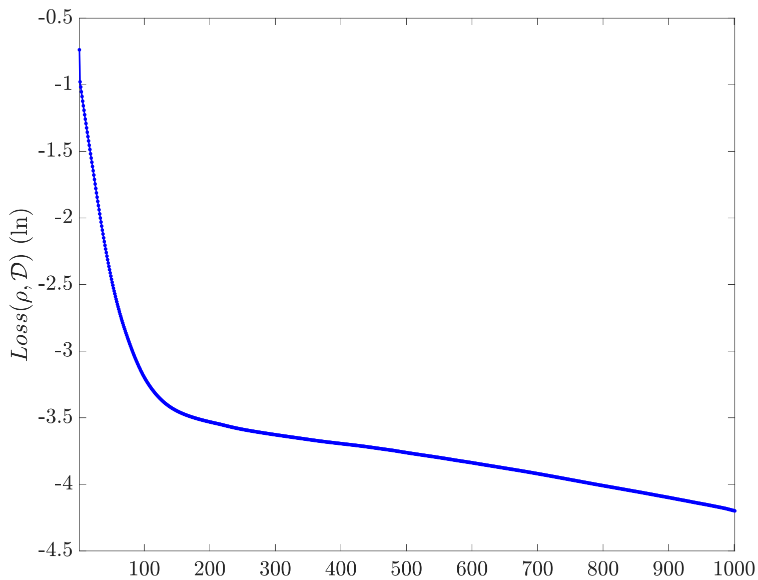


Figure 11.3.1: Loss function $Loss(\boldsymbol{\rho}, \mathcal{D})$ across simulations.

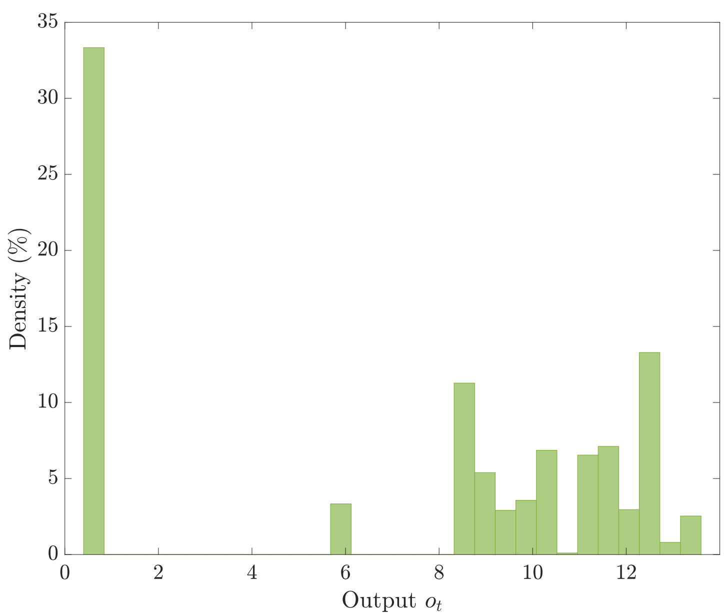


Figure 11.3.2: Probability distribution of output

A Mathematical Appendix

Some of the basic mathematics used in the book is reviewed here. This should make the book self-contained for those rusty or unfamiliar with the mathematics used. The presentation is cookbook in style and is oriented toward discussing the uses of mathematics in the main text. An excellent gentle and gradual introduction to the mathematics used in economics is contained in Chiang (2011). A good introduction to probability and statistics is DeGroot (1975). Last, Bryant (1985) provides a heuristic approach to real analysis.

A.1 Notation

- Upper case Roman letters usually denote a function.
- Lower case Roman letters usually represent a variable.
- Calligraphic letters are often used to represent sets or spaces.
- Greek letters are usually parameters.
- $\equiv$ is a symbol meaning equal to by definition.
- $\simeq$ denotes approximately equal to.
- $\in$ is shorthand for contained in.
- $|x|$ is either the absolute value of the scalar x or the norm of a vector x.
- $\mathcal{R}$ represents the real numbers and $\mathcal{R}_+$ is the positive reals.
- $F : \mathcal{X} \to \mathcal{Y}$. The function F maps the space $\mathcal{X}$ (the domain) into the space $\mathcal{Y}$ (the range).
- $F(x, y)$ denotes a function of two variables, x and y.
- F_1, F_2, F_{11}, F_{12}, and F_{22}. These denote various derivatives of the function F. Specifically, F_1 is the partial derivative of F with respect to its argument, x. Thus, $F_1 \equiv dF/dx$. Likewise, $F_2 \equiv dF/dy$. Next, F_{11} is the derivative with respect to x of the first derivative F_1 so that $F_{11} \equiv d^2F/dx^2$. Finally, $F_{12} \equiv d^2F/(dxdy)$ and $F_{22} \equiv d^2F/dy^2$.

- e or exp is Euler's constant or $2.7182\cdots$.
- ln is the natural logarithm. I.e., the logarithm using the base e.
- Pr is shorthand for probability.
- $\mod$ is the modulo operator. Used to characterize the remainder from a division.
- FLOOR is notation for round down to the nearest natural number.
- i is the square root of -1, which is an imaginary number.
- E is the expectations operator. So, $E[x]$ is the expected value of x.
- $'$ used to signify the value of a variable one period down the road.

A.2 Maximizing a Function

Maximization is at the heart of economics. Economic actors try to do the best for themselves. So, people maximize their utility and firms maximize their profits. Mathematically speaking this corresponds to maximizing a function. Functions are everywhere in economics: cost functions, production functions, utility functions to name a few.

Consider the function

$$y = F(x),$$

which maps the real-valued variable x into a real value for the variable y. By definition, a function associates each value of x with a *unique* value for y. Take x to be a nonnegative number, so $x \in \mathcal{R}_+$, and y to be some real number, implying $y \in \mathcal{R}$. Thus, $F : \mathcal{R}_+ \to \mathcal{R}$. Assume that F is continuously twice differentiable. Denote the first and second derivatives of F by

$$F_1(x) \equiv \frac{dF(x)}{dx} \text{ and } F_{11}(x) \equiv \frac{d^2F(x)}{dx^2} = \frac{dF_1(x)}{dx}.$$

The first derivative gives the impact that a small change in x will have on y. The second derivative specifies how the first derivative changes in response to a small shift in x. In other words, it says how the change in y in response to a tiny shift in x, itself, changes with a small movement in x.

Now, consider the unconstrained maximization problem

$$\max_x F(x). \tag{A.2.1}$$

Here the value of x is sought that maximizes the function $F(x)$. At a maximum, the following first-order condition must obtain

$$F_1(x) = 0.$$

This condition is necessary for a local maximum. Suppose to the contrary that at a maximum $F_1(x) > 0$. Then, a small shift up in x would increase $F(x)$, a contradiction. The above first-order condition represents one equation in one unknown, x. The first-order condition specifies a local maximum, instead of a local minimum (or an inflection point), if the second-order condition shown below holds

$$F_{11}(x) < 0.$$

Let x^* denote the value of x that maximizes the the function, $F(x)$. When the second-order condition holds, a small increase in x must cause the function $F(x)$, when evaluated at x^*, to decrease, because $F_1(x)$ becomes negative. Likewise, a small decrease in x induces $F_1(x)$ to become positive, implying that the reduction in x also results in a decline in $F(x)$. Therefore, x^* must maximize $F(x)$, at least locally.

A.2.1 *Strict Concavity (Convexity) and the Second-Order Condition for a Maximum (Minimum)*

Now, a strictly concave function has a negative second derivative. That is, if a function is strictly concave, then $F_{11}(x) < 0$ for all x. In this situation, the second-order condition for a maximum will automatically hold; hence, for strictly concave functions the first-order condition is both necessary and sufficient for characterizing a maximum. By contrast, a strictly convex function has a positive second derivative so that $F_{11}(x) > 0$ for all x. In this case, the first-order condition is both necessary and sufficient for a minimum to hold.

This situation is portrayed by Figure A.2.1 for a typical case in economics. At the peak of the function the slope or the first-derivative is zero. Note that the second-derivative is negative. That is, the first-derivative declines as you move from left to right. This occurs because the objective function is strictly concave in x.

A.2.2 *Envelope Theorem*

Rewrite problem (A.2.1) as

$$V(\alpha) = \max_x F(x;\alpha),$$

where α is an exogenous parameter. The function $V(\alpha)$ gives *optimized* value of F for the a given value of α. One can ask how this optimized value of F changes with the parameter α. Now, let x^* be the optimal value for x. This value for x must solve the first-order condition attached to the above problem. That is, x^* must solve the equation

$$F_1(x^*;\alpha) = 0.$$

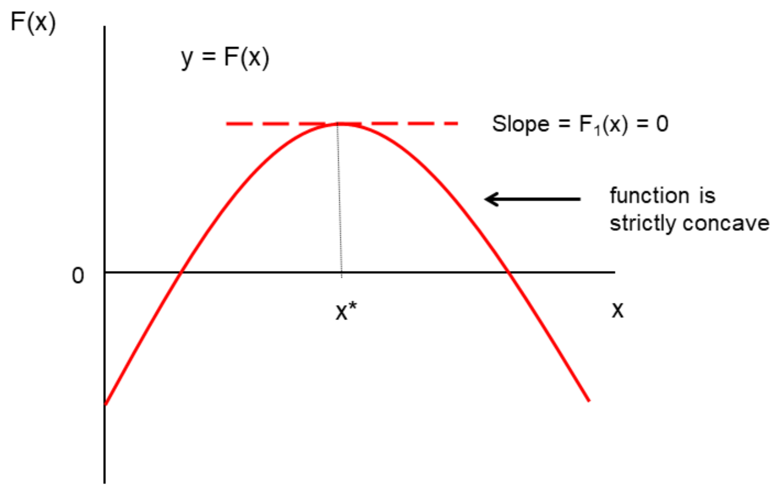


Figure A.2.1: Finding an unconstrained maximum.

By definition,

$$V(\alpha) = F(x^*; \alpha),$$

because x^* is the value of x that maximizes $F(x)$. Differentiate both sides of the above expression with respect to α to get

$$\frac{dV(\alpha)}{d\alpha} = \underbrace{F_1(x^*; \alpha)}_{=0} \frac{dx^*}{d\alpha} + F_2(x^*; \alpha).$$

By the first-order condition, $F_1(x^*; \alpha) = 0$, any induced variation in x^* caused by a change in α washes out implying

$$\frac{dV(\alpha)}{d\alpha} = F_2(x^*; \alpha).$$

This occurs because at the maximum a small change in x will have no impact on the objective function. As can be seen from Figure A.2.1, at the top of the function an infinitesimal step left or right won't change the value of the function. The envelope theorem is used in Chapters 6 and 9.

A.2.3 Corner Solutions

In economics corner solutions to maximization problems often occur. For example, perhaps a person wants to set their hours worked in the labor force to be zero, or likewise, they do not want to acquire any skill by attaining a post-secondary education. Now, suppose that there is a lower bound on x, denoted by x_l, so that the constraint $x \geq x_l$ must hold. The maximization problem above now appears as

$$\max_{x \geq x_l} F(x).$$

One of two solutions may obtain to this constrained maximization problem; viz., an interior solution or a corner solution. The interior solution is described as before by the first-order condition

$$F_1(x) = 0.$$

The corner solution occurs when

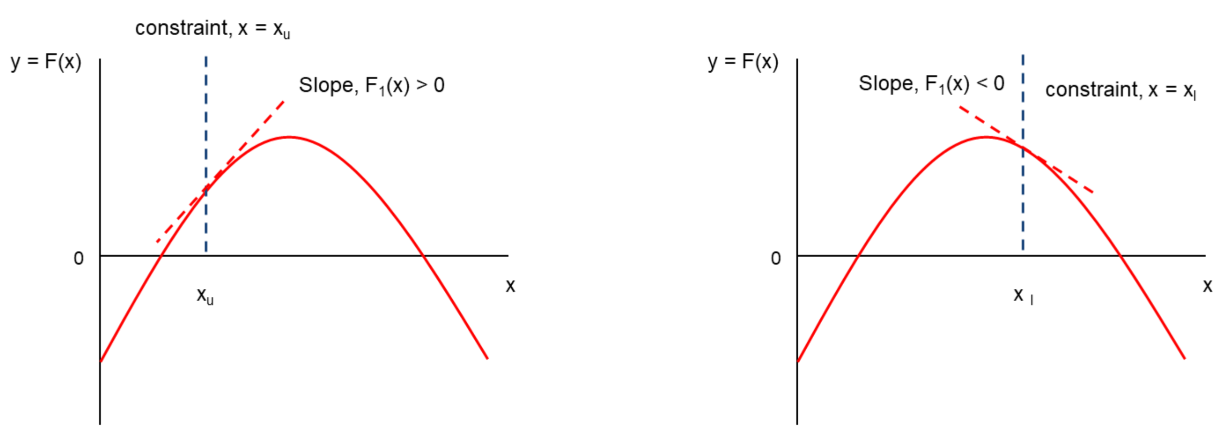
$$F_1(x_l) < 0.$$

This is shown by the right-hand panel of Figure A.2.2. Here, the peak of the function cannot be attained because the lower bound on x has been hit. The slope is negative at $x = x_l$. Because $F_1(x) < 0$ at $x = x_l$, a small reduction in x would increase the value of the objective function, $F(x)$. This cannot be done due to the presence of the lower bound, x_l. Alternatively, x could be constrained by an upper bound, x_u, requiring that $x \leq x_u$. Now the corner solution happens when

$$F_1(x_u) > 0.$$

A small increase in x from x_u would raise the value of objective function, but this isn't feasible, because the upper bound, x_u, has been hit. The left-hand side of Figure A.2.2 illustrates this situation.

Figure A.2.2: Constrained maximization. The right-hand side panel shows the situation when a corner solution is hit at a lower bound, while the left-hand side illustrates things for an upper bound.



A.2.4 Constrained Maximization

Optimization in economics often involves maximization subject to constraints. For example, a consumer maximizes utility subject to a budget constraint or a firm minimizes costs subject to a production function. Consider maximizing the function $F(x, y)$, with respect to the decision variables x and y, subject to a constraint given by the function $y = G(x)$. There are two ways to proceed here. First, one could just replace y in the objective function with the function $G(x)$ and then maximize with respect to the single variable, x. That is, one could solve the problem

$$\max_x F(x, G(x)).$$

By defining the new function $\widetilde{F}(x) = F(x, G(x))$, it should be clear that this reduces to the form of the maximization problem discussed above.

A.3 Total Differentials

Consider the function

$$z = F(x, y).$$

What would happen to z if both x and y are changed by some arbitrary small amounts? Denote the small changes in x and y by dx and dy, respectively. These are called differentials. Likewise, the induced total change in z is represented by dz. The total change in z is given by

$$dz = F_1(x, y)dx + F_2(x, y)dy, \tag{A.3.1}$$

where $F_1(x, y) \equiv dF(x, y)/dx$ and $F_2(x, y) \equiv dF(x, y)/dy$. The above expression decomposes the change in z into two factors. The first term on the right-hand side is the change in z that results from the shift in x. The shift in x is represented by dx. To get the induced shift in z, this is multiplied by the (partial) derivative $F_1(x, y)$, which translates a shift in x into a shift in z. The second term does the same thing for y.

A.3.1 The Total Derivative

The differentials dz, dx, and dy can be manipulated to obtain derivatives. For example, one could divide the above equation through by dx to obtain

$$\frac{dz}{dx} = F_1(x, y) + F_2(x, y)\frac{dy}{dx}.$$

The term dz/dx is the total derivative of z with respect to x. The change in x has both a direct and indirect effect on z, as shown by the first and second terms on the right-hand side. The indirect effect occurs because the change in x may induce a change in y, as given by dy/dx, which in turn will affect z via $F_2(x, y)$. To compute the indirect effect more information would be needed. To illustrate, perhaps y is given by the function $y = G(x)$. Then, $dy/dx = G_1(x)$, implying $dz/dx = F_1(x,y) + F_2(x,y)G_1(x)$. Alternatively, perhaps y is not a function of x. Then, $dy/dx = F_1(x,y)$. As another example, perhaps x and y are functions of another variable t, which changed by the small amount, dt. Then, dividing both sides of (A.3.1) by dt gives

$$\frac{dz}{dt} = F_1(x,y)\frac{dx}{dt} + F_2(x,y)\frac{dy}{dt}.$$

Here, the derivatives dx/dt and dy/dt depend on the specified functional dependencies of x and y on t.

A.4 Intermediate Value Theorem

Let $F(x)$ be a continuous function whose domain contains the interval $[a,b]$. The function $F(x)$ takes on every value between $F(a)$ and $F(b)$ as x traverses the interval $[a,b]$.

A.5 The Implicit Function Theorem

Consider an equation of the form

$$F(x,y) = 0.$$

Here F is a function. Think about x as being an endogenous variable and y as being an exogenous one. Then, the above expression represents one equation in one unknown variable, x. Does a solution exist where one could write

$$x = G(y), \tag{A.5.1}$$

where G is some function? Now, let $F : \mathcal{R}^2 \to \mathcal{R}$ be a continuously differentiable function. The implicit function theorem states that, provided $F_1(x,y) \neq 0$ at the solution point, there will indeed be a continuously differentiable solution of the form (A.5.1) where

$$F(G(y),y) = 0.$$

A.6 First- and Second-Order Taylor Expansions

First- and second-order Taylor expansions, without the remainder terms, are illustrated for the case of a bivariate function. Let $F(x,y)$ be a twice differentiable function of two variables, x and y. The function $F(x,y)$ can be approximated around the point (x^*,y^*) by using either a first- or second-order Taylor expansion as follows:

$$F(x,y) \simeq F(x^*,y^*) + F_1(x^*,y^*)(x-x^*) + F_2(x^*,y^*)(y-y^*), \quad \text{(first order)}$$

and

$$F(x,y) \simeq F(x^*,y^*) + F_1(x^*,y^*)(x-x^*) + F_2(x^*,y^*)(y-y^*) + \frac{1}{2}F_{11}(x^*,y^*)(x-x^*)^2 + F_{21}(x^*,y^*)(x-x^*)(y-y^*) + \frac{1}{2}F_{22}(x^*,y^*)(y-y^*)^2, \text{ (second order).}$$

A.7 *Unimodal Function*

Definition A.1. (Unimodal function) A function $F(x)$ is unimodal if for some value x^*, it is monotonically increasing (decreasing) for $x \leq x^*$ and monotonically decreasing (increasing) for $x \geq x^*$. Clearly, the only maximum (minimum) for $F(x)$ is $F(x^*)$.

A.8 *The Golden Ratio*

The golden ratio, often denoted by ψ is the positive solution to the polynomial $\psi^2 - \psi - 1 = 0$. By using the quadratic formula, it is easy to calculate that $\psi = (1+\sqrt{5})/2 = 1.61803398874\cdots$. Interestingly, $1/\psi = 0.61803398874\cdots$, because the above polynomial can be restated as $\psi - 1 = 1/\psi$.

A.9 *Homogenous Function*

Definition A.2. (Homogenous function) Let $x = (x_1, x_2, \cdots, x_n)$ be a n-dimensional vector. A function $y = F(x)$ is said to be homogenous of degree q if for all x,

$$F(\lambda x) = \lambda^q F(x), \text{ for } \lambda \neq 0.$$

In otherwords, if the all of the arguments in the function F, or $x_1, x_2, \cdots, x_n$, are scaled by the common factor λ, then the output of the function, y, rises by the factor λ^q.

A.10 *Euler's Theorem*

Lemma A.1. *(Euler's theorem) Consider a function, $F(k,h)$, which is homogenous of degree one in k and h; i.e., exhibits constant returns to scale in k and h. Then,*

$$F(k,h) = F_1(k,h)k + F_2(k,h)h.$$

Proof. Since F is homogenous of degree one in k and h,

$$\lambda F(k,h) = F(\lambda k, \lambda h).$$

Differentiating with respect to λ then gives

$$F(k,h) = F_1(k,h)k + F_2(k,h)h.$$

□

Remark A.1. In the main text, $F(k,h)$ is a constant-returns-to-scale production function and $F_1(k,h)$ and $F_2(k,h)$ are the marginal products of capital and labor. In competitive equilibrium $F_1(k,h)$ and $F_2(k,h)$ will be equal to the rental rate on capital, r, and the wage wage, w. Therefore, $F(k,h) = rk + wh$.

A.11 Eigenvalues and Eigenvectors

Let T be a $n \times n$ matrix. An eigenvalue/eigenvector pair satisfy the equation

$$eT = \varepsilon e,$$

where e is the $1 \times n$ (left) eigenvector and the scalar ε is the associated eigenvalue. This can also be expressed as

$$Te' = \varepsilon e',$$

where e' is the $n \times 1$ (right) eigenvector and again with the scalar ε being the associated eigenvalue. As is probably obvious, is e' is just the transpose of e.

The eigenvalues of the matrix T solve the equation

$$\det(T - \varepsilon) = 0.$$

This equation yields a polynomial of degree n that may have up to n distinct (potentially complex) roots or eigenvalues. The eigenvalues are the n values of ε that solve the characteristic polynomial

$$\det(T - \varepsilon) = (\varepsilon - v_1)(\varepsilon - v_2) \cdots (\varepsilon - v_n) = 0.$$

These values may be repeated and complex.

A.12 Descriptive Statistics

A.12.1 Mean and Median

The mean is just the average value of the data in a set. When the data is ordered from the lowest to the highest value, the median is the middle value.

Definition A.3. (Mean) Let $\{x_i\}_{i=1}^{N}$ be a data series. The mean for the series, μ_x, is defined by

$$\mu_x = \frac{1}{N} \sum_{i=1}^{N} x_i.$$

Definition A.4. (Median) Let $\{x_i\}_{i=1}^N$ be a data series ordered without loss of generality such that $x_i \geq x_{i-1}$ for $2 \leq i \leq N$. The median for the series, MEDIAN_x, is defined by

$$\text{MEDIAN}_x = \begin{cases} x_{(N+1)}, & \text{for } N \text{ odd;} \\ (x_{N/2} + x_{N/2+1})/2, & \text{for } N \text{ even.} \end{cases}$$

A.12.2 *Standard Deviation*

The standard deviation measures the amount of dispersion or variation in a data series. The bigger the number is the higher is the amount of dispersion around the series's mean. In business cycle analysis the standard deviation of a series measures its fluctuations over time.

Definition A.5. (Standard Deviation) Let $\{x_i\}_{i=1}^N$ be a data series. The standard deviation for the series, σ_x, is defined by

$$\sigma_x = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu_x)^2} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}x_i^2 - \mu_x^2},$$

where the mean, μ_x, is

$$\mu_x = \frac{1}{N}\sum_{i=1}^{N} x_i.$$

A.12.3 *Pearson Correlation Coefficient*

Correlation coefficients measure the association between two data series, say $\{x_i\}_{i=1}^n$ and $\{y_i\}_{i=1}^n$. A correlation coefficient takes a value between -1 and 1, where a positive value indicates that two series tend to move together while a negative one shows that they have a proclivity to move oppositely to one another. The higher the correlation coefficient is in absolute value the stronger is the association. A value of 0 shows no association. The Pearson correlation coefficient measures the degree of linear association between two series. In business cycle analysis often one is interested in how a variable moves with GDP. When a variable has a positive correlation (negative correlation) with GDP it is called procyclical (countercyclical).

Definition A.6. (Pearson correlation coefficient) The Pearson correlation coefficient is a measure of the linear dependence (or correlation) between the two data series, $\{x_i\}_{i=1}^n$ and $\{y_i\}_{i=1}^n$. It is defined by the formula

$$\rho = \frac{\sum_i^n (x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_i^n (x_i - \overline{x})^2}\sqrt{\sum_i^n (y_i - \overline{y})^2}},$$

where the sample means, $\overline{x}$ and $\overline{y}$, are given by $\overline{x} = (\sum_i^n x_i)/n$ and $\overline{y} = (\sum_i^n y_i)/n$. If there is a tendency when x rises above its mean for y to do so as well, then the numerator will likely be positive and hence so will be ρ. The opposite will be true, if there is a penchant for y to fall below its mean when x rises above its one. If y and x have a strictly positive (negative) linear relationship then ρ will be 1 (-1).

Figure A.12.1 illustrates the Pearson correlation coefficient between x and y for some randomly generated series. The two series for x and y are always positively associated. As can be seen, as ρ increases so does the strength of the positive association.

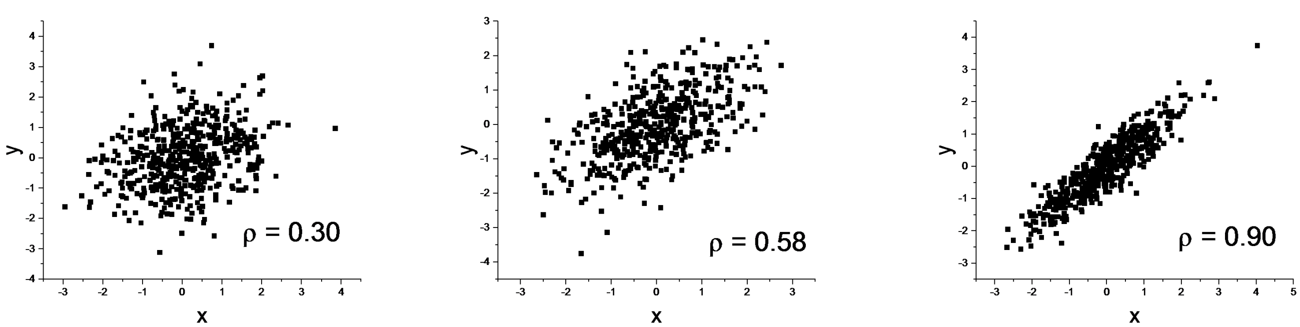


Figure A.12.1: Pearson correlation coefficient. As one moves from left to right, the degree of positive association between the series for x and y increases. This is reflected in higher values for the Pearson correlation coefficient, ρ.

A.12.4 *Coefficient of Autocorrelation*

The autocorrelation of a time series measures the correlation of the series with a delayed facsimile of itself. In business cycle analysis it is used to measure the degree of persistence in a time series.

Definition A.7. (Autocorrelation) The autocorrelation coefficient is just the correlation coefficient between the current and lagged value of a variable.

The autocorrelation coefficient, since it is just a correlation coefficient, has a value between -1 and 1.

A.13 *The Uniform, Normal, and Weibull Distributions*

Definition A.8. (Uniform Distribution) A random variable $\tilde{x}$ is distributed according to a uniform distribution $U : [\underline{x}, \overline{x}] \to [0,1]$ if

$$\Pr[\tilde{x} \leq x] = U(x) = \frac{x - \underline{x}}{\overline{x} - \underline{x}},$$

for $\underline{x} \leq x \leq \overline{x}$. The uniform distribution $U(x)$ is just a straight line that starts at 0 when $x = \underline{x}$ and ends at 1 when $x = \overline{x}$. The probability density function connected with the uniform distribution is

$$U_1(x) = \frac{1}{\overline{x} - \underline{x}}.$$

The mean and variance of x are given by $(\overline{x} + \underline{x})/2$ and $(\overline{x} - \underline{x})^2/12$.

Definition A.9. (Normal Distribution) A random variable $\tilde{x}$ is distributed according to a normal distribution $N(\mu, \sigma^2) : (-\infty, \infty) \to [0,1]$, with mean μ and variance σ^2, if

$$\Pr[\tilde{x} \leq x] = \int_{-\infty}^{x} \frac{1}{\sigma(2\pi)^{1/2}} \exp[\frac{-(\tilde{x} - \mu)^2}{2\sigma^2}] d\tilde{x},$$

where

$$\frac{1}{\sigma(2\pi)^{1/2}} \exp[\frac{-(\tilde{x} - \mu)^2}{2\sigma^2}],$$

is the probability density function for a normal distribution. When the logarithm of a variable is normally distributed the variable is said to follow a log-normal distribution.

Definition A.10. (Bivariate Normal Distribution) Two random variables $\tilde{x}$ and $\tilde{y}$ are distributed according to a bivariate normal distribution N with means μ_x and μ_x, variances σ_x^2 and σ_y^2, and correlation ρ if where

$$\frac{1}{2\pi\sigma_x\sigma_y\sqrt{(1-\rho^2)}} \times \exp\{-\frac{1}{2(1-\rho^2)}[\frac{(\tilde{x} - \mu_x)^2}{\sigma_x^2} - 2\rho\frac{(\tilde{x} - \mu_x)(\tilde{y} - \mu_y)}{\sigma_x\sigma_y} + \frac{(\tilde{y} - \mu_y)^2}{\sigma_y^2}]\},$$

is the probability density function for a normal distribution. When the logarithm of a variable is normally distributed the variable is said to follow a log-normal distribution.

Definition A.11. (Weibull Distribution) A random variable $\tilde{x}$ is distributed according to a Weibull distribution $W : [0, \infty) \to [0,1]$ if

$$\Pr[\tilde{x} \leq x] = W(x) = 1 - \exp[-(x/\eta)^\beta],$$

Here $\eta > 0$ is called the scale parameter and $\beta > 0$ is referred to as the shape parameter. Depending on parameter values, the density function for the Weibull function can fall or rise and then fall. The Weibull distribution has an easy formula for the median of the distribution:

$$\eta \ln(2)^{1/\beta}.$$

The formulae for the mean and variance are somewhat more complicated:

$$\eta\Gamma((1+\beta)/\beta)$$

and

$$\eta^2\Gamma((2+\beta)/\beta) - \eta\Gamma((1+\beta)/\beta)^2,$$

where Γ is the gamma function. The gamma function is built into most numerical programming languages, such as MATLAB.

A.14 *The Strong Law of Large Numbers*

Let $\{x_i\}_{i=1}^n$ be a sample of independently and identically distributed random numbers drawn from some distribution with mean μ. Let the mean of the random sample be denoted by

$$x_n = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

The strong law of large numbers states that

$$\Pr[\lim_{n\to\infty} x_n = \mu] = 1.$$

In other words, as the sample sizes increases the mean of the sample will approach the mean of the distribution with virtual certainty.

B Introduction to MATLAB

This tutorial presents an introduction to MATLAB.[1] The primary programming language used here is MATLAB, a powerful and popular language to solve numerical problems, including numerical integration, numerical optimization, random number generation, and simulation. To embark on this programming adventure, this tutorial covers five basic topics as outlined on the next pages: getting started, basic commands, flow control, defining functions, and graphing. The tuturorial also explains how to solve a nonlinear equation.

As in learning any programming language, making progress in programming skill is –as always– *"learning by doing"*. From this perspective, the following sections of this tutorial will only serve as the key to open the door of MATLAB, beyond which there is a vast universe for you to explore.

[1] The core of this tutorial was written by Pengfei Han.

B.1 Getting Started

B.1.1 Interface of MATLAB

Once clicking on the icon of "MATLAB" on your computer, there will be five windows popping up on the screen:

Current Directory:
This is where your current MATLAB files are stored and managed.

Command Window:
This is where you type in your commands to MATLAB.

Command History:
This window tracks a sequence of your recent commands.

Workspace:
This is where the variables in your current program are kept.

Variable Editor:
The contents of the variables can be examined and edited in this section.

B.1.2 *Housekeeping*

Prior to the main body of the coding work, there are three basic commands commonly used for housekeeping:
CLEAR ALL: removes all variables from the current workspace.
CLOSE ALL: deletes all figures whose handles are not hidden.
CLC: clears all input and output from the command window, giving you a "clean screen".

B.1.3 *Loading Data*

To import data into MATLAB, simply use the import wizard:
"HOME" $\rightarrow$ "Import Data", and choose the folder/file you wish to import.
Alternatively, load the data by using the commands associated with the type of file to be imported. For instance:
to import files with format "xls", use the command: xlsread('FileName.xls');
to import files with format "csv", use the command: csvread('FileName.csv');
A detailed example is given in the M-file "Ex_load_data.m".

B.1.4 *Writing/Reading Results to a Table or Matrix*

It's easy to write results to a table in MATLAB using the TABLE command. Suppose one has two $n \times 1$ vectors x and y. These can be written to a table called T using the following command:

$$T = \text{table}(x, y);$$

This will create a $n \times 2$ table with the headings x and y for each column. This table can be exported to an EXCEL file called Data using the WRITETABLE command. Specifically, write

filename = 'Data.xlsx'

and

writetable(T, filename);

Similarly, it is easy to write results to a matrix using the WRITEMATRIX command; i.e., WRITEMATRIX(matrix,filename). These commands are used in Chapter 2 for the Lucas (1987) welfare calculations concerning changes in an economy's growth rate.

The analogous commands for reading tables and matrices into MATLAB are READTABLE and READMATRIX.

B.1.5 *Stopwatch Timer*

To monitor the performance of our code, a stopwatch timer can be used to measure the execution time. This stopwatch timer starts with the TIC at the begging of the code, and to display the elapsed time, simply use the command TOC.

B.1.6 *Help*

Whenever you get lost in your programming endeavor, always recall the most powerful command in MATLAB: HELP. Help can be obtained two ways. First, you can access the help menu. Just click on the ? on the upper righthand side of the screen. Then navigate your way through the index until you see what you want. Second, if you know the name of the command or function that you are interested in, then in the command window just type HELP name. This is good for seeing the syntax associated with executing a command or the options that are available.

B.1.7 *A Line in a MATLAB File*

A line in a MATLAB file is usually an executable statement. Usually it ends with a semicolon or ;. This tells MATLAB to run the line silently; i.e., not to print the result of the executable statement to the screen. If you want to see the output, then omit the semi colon. Often your executable statement will run over one line. Then, you must put a continuation statement at the end of line in the Editor before you go onto the next line. This way MATLAB knows that the next line is part of the same executable statement. The continuation statement is just three dots or . . . .

B.2 *Basic Commands*

As revealed in the term “MAT" as in MATLAB, this language is particularly optimized in processing matrices, so as the first programming suggestion, it is always a weakly dominating strategy to write your code in a “matrix" (in contrast to a “loop"), whenever you can. The codes of the examples in this section can be found in the M-file “**Ex_basic_command.m**".

B.2.1 *Creation and Concatenation of Matrices*

Creation of Matrices

Creating A General Matrix
The matrix is the central element for MATLAB. In general you can simply create a matrix by enumerating its elements as follows:
M = [1 2 5; -1 20 7; 8 -9 3]
This creates a 3 by 3 matrix: M.

Identity, Zero, and One Matrix
Very frequently you will need to create objects such as the identity matrix, matrices of zero, and matrices of one. MATLAB has handy functions for all of these tasks:
I = EYE(5, 5)
This creates a 5-by-5 identity matrix.
O = ZEROS(2, 3)
This creates a 2-by-3 matrix with elements of zero: O.
Y = ONES(3, 4)
This creates a 3-by-4 matrix with elements of one: Y.
Creating a Vector
Also frequently used in MATLAB are vectors, which can be created this way:
V = 0 : 0.5 : 10;
This creates a row vector V with elements ranging from 0 to 10 with increments of 0.5.

Concatenation of Matrices

In MATLAB the concatenation operator is “[ ]".
For example, you can join two matrices side by side as follows:
X = zeros(2, 2);
Y = ones(2, 2);
join = [X Y]
0 0 1 1
0 0 1 1
To stack the two matrices, use the command “[;]":
stack = [X; Y]
0 0
0 0
1 1
1 1
To delete the jth column from X write X(:,j) = [];
Analogously, to delete the ith row write X(i,:) = [];

B.2.2 *Operators and Relational Operators*

Operators

There are four basic operators in MATLAB:
+ : Addition
- : Subtraction
* : Multiplication (applicable for both scalars and matrices)
∧ : Matrix Power
In addition, there are three element-by-element operators:
.* : Multiplication, operated element by element

./ : Division, operated element by element
.∧ : Power, operated element by element

Moreover, listed as follows are some frequently used operators for your reference:
EXP(X): the e-exponential of the elements of the matrix X
LOG(X): the natural logarithm of the elements of the matrix X
SQRT(X): the square root of the elements of the matrix X
ABS(X): the absolute value of the elements of the matrix X
ROUND(X) - rounds the elements of X to the nearest integers
FLOOR(X): rounds the elements of X to the nearest integers towards minus infinity
CEIL(X): rounds the elements of X to the nearest integers towards infinity
DIAG(X): the diagonal elements of XDIAG(X): the diagonal elements of X
EPS(X): Returns floating-point relative accuracy. In particular, it returns the positive distance from ABS(X) to the next larger floating-point number of the same precision as X. If X is omitted, it is assumed to be one.
FIX(X) : rounds the elements

Relational Operators

There are eight basic relational operators in MATLAB:
< : less than
> : greater than
<= : less than or equal to
>= : greater than or equal to
== : equal to
∼= : not equal
&& : and
||: or

B.2.3 *Matrix Commands*

DET, INV, MAX, MIN, MEAN, NORM, PROD, RANK, SUM

MAX(X): returns the maximum values of matrix X along its columns. This command is extremely useful for discrete-state-space dynamic programming, as discussed in Chapter 9.
For instance: X = [1 2 3; 4 5 6; 7 8 9]
1 2 3
4 5 6
7 8 9

MAX(X)
7 8 9
One may also want to retrieve the location of the maximal elements in addition to the maximun values. To do this, write [value,location] = MAX(X). To obtain the maximum along the rows, use the command: max(X, [ ], 2). The "2" indicates that you are interested in the second dimension of the matrix. For instance:
MAX(X, [ ], 2)
3
6
9
Additonally, [value, location] = MAX$([x_1, x_2, \cdots, x_n])$ returns the maximum element and its location in the vector $[x_1, x_2, \cdots, x_n]$. This command is useful for solving discrete-state-space dynamic programming problems as discussed in Chapter 9. You can also apply this operator on matrices. For instructions on how to do this look at the help menu in MATLAB. Analogously, apply the matrix operators: MIN to find the minimum, MEAN to obtain the mean, and PROD to calculate the products. For instance: X = [1 2 3; 4 5 6; 7 8 9]
1 2 3
4 5 6
7 8 9
MEAN(X)
4 5 6
MEAN(X, 2)
2
5
8
SUM(X)
12 15 18
SUM(X, 2)
6
15
24
PROD(X)
28 80 162
PROD(X, 2)
6
120
504
NORM(X) returns the Euclidean norm (or the square root of the sum of the squares) for the matrix X. This command is often used to check whether an algorithm has converged.
INV(X) returns the inverse of X, when X is a square matrix. This com-

mand can be used to find the multiplier matrix associated with the Leontief input-output matrix–see Chapter 2
DET(X) is the the determinant of X.
RANK(X) is the rank of X.
ISNAN(X) returns an matrix of ones and zeros where a one denotes that an element of X is not a number (NaN), perhaps because its value is missing, a character, has an infinite value.
FIND(x): returns a vector containing the linear indices of elements greater than zero
FIND(x>a): returns a vector containing the linear indices of elements greater than a. FIND is a useful command. For example, perhaps you loaded in a matrix of data from somewhere. Some entries might be missing, a character, or have an infinite value. You can search for these entries by writing FIND(isnan(matrix)).

Obtain Dimensions of A Matrix

[M, N] = SIZE(X): returns the number of rows (M) and columns (N) of the matrix X as separate output variables.
SIZE(X, 1): returns the number of rows (M)
SIZE(X, 2): returns the number of columns (N)
LENGTH(X) : returns the length of the vector X

B.3 *Flow Control*

Structural command – or flow control – governs the flow of information in the program and, thus, is essential in any programming language. In particular, there are three structural commands in MATLAB: IF, FOR, and WHILE. The codes associated with this section are in the M-file "EX_FLOW_CONTROL.M".

B.3.1 IF

The "IF" statement –matched with the command "END" – executes a group of statements when a logical expression is evaluated to be true. The general form is:
IF logical expression A
statements to be executed if A is true
END
In addition, the logical evaluation process can be enriched by adding the command "ELSE":
IF logical expression A
statements to be executed if A is true
ELSE

statements to be executed if A is false
END
Moreover, the if-statement can be augmented by adding a series of the command "ELSEIF":
IF logical expression A1
statements to be executed if A1 is true
ELSEIF logical expression A2
statements to be executed if A2 is true
ELSE
statements to be executed if all logical expressions enumerated above are false
END

IF statements are very useful. For example, perhaps you are solving a nonlinear equation. You could ask the computer to display a message if the zero for the equation exceeds some desired level of tolerance.

Application: How Much Do You Pay for Federal Income Tax?

How much is an average American family paying for the Federal income tax?
Given the tax rate schedule, the if command can deliver the federal income tax for any level of income:

```
% median household income (current $) in the United States: $53,657 in 2014
income = 54462;
if income > 0 && income <= 12950
tax = income * 10 * .01;
elseif income > 12950 && income <= 49400
tax = 1295.00 + (income - 12950) * 15 * .01;
elseif income > 49400 && income <= 127550
tax = 6762.50 + (income - 49400) * 25 * .01;
elseif income > 127550 && income <= 206600
tax = 26300.00 + (income - 127550) * 28 * .01;
elseif income > 206600 && income <= 405100
tax = 48434.00 + (income - 206600) * 33 * .01;
elseif income > 405100 && income <= 432200
tax = 113939.00 + (income - 405100) * 35 * .01;
else % i.e., income > 432200
tax = 123424.00 + (income - 432200) * 39.6 * .01;
end
```

Rate Schedule for the Federal Income Tax (2014)			
Schedule Z (Applies to the Head of Household)			
If Taxable Income Is Over	*But Not Over*	*The Tax Is:*	*Of the Amount Over*
0	12,950	10%	0
12,950	49,400	1,295.00 + 15%	12,950
49,400	127,550	6,762.50 + 25%	49,400
127,550	206,600	26,300.00 + 28%	127,550
206,600	405,100	48,434.00 + 33%	206,600
405,100	432,200	113,939.00 + 35%	405,100
432,200	∞	123,424.00 + 39.6%	432,200

B.3.2 For

This is a very important command. The FOR loop repeats a group of statements by a fixed and predetermined number of times. The general form of the for loop is:

```
FOR n = 1 : N
statements to be repeated
END
```

In this loop, the variable "n" – which begins at 1 and ends at N – serves as a counter, and the variable "N" controls the number of repetition. For example, one may want to iterate down a time path for variable. In this case n would be a period. Or perhaps one wants to do some sort of operation across a group of individuals. In the case n would be a person.

Application: How Fast Is the Federal Deficit Growing?

How fast is the deficit of the federal government growing?
Given the time series of federal deficit obtained in Section 1.3 ("loading data"), this task is straightforward to tackle by a for loop:

```
for id = 1 : (n_year - 1)
g_def(id) = GDP(id + 1) / GDP(id) - 1;
end
```

B.3.3 While

The WHILE loop repeats a group of statements an indefinite number of times, under the control of a logical condition. This is a very useful command when writing an algorithm. One may want an algorithm to keep on going while some convergence criteria hasn't been met or while some maximum number of iterations hasn't been exceeded.
The general form of the while loop is:

WHILE logical expression
statements to be repeated
END

Application: How to Find the Steady State In Solow Growth Model

Consider a Solow growth model in which the capital is accumulated by:

$$k_{t+1} = s\,k_t^{\alpha} + (1-\delta)k_t$$

What is the steady state level of capital in this economy?
This problem can be readily solved by the while command:

```
% Parameterization
s = 0.0550; % savings rate
aalpha = 1/3; % capital income share
ddelta = 0.145; % depreciation rate of capital

% Initialization
diff = 1; % specify the initial distance from convergence
criterion = 1e-8; % the criterion to determine convergence
k_new = 1; % initial guess for the steady-state capital

while diff > criterion
   % characterize the evolution of capital with exogenous savings rate
   k_old = k_new;
   k_new = s * (k_old)(aalpha) + (1 - ddelta) * k_old;
   % evaluate the distance from converging to the steady state
   diff = abs(k_new - k_old);
end
```

B.4 *Defining Functions*

B.4.1 *Anonymous Function*

Functions are used extensively in numerical work. The most basic form of the function is an anonymous function and can be simply created by the command @() in MATLAB.
For instance: cubic = $@(x)x \wedge 3$;
This creates a function "cubic" which transforms any input to its power of 3.
To call this function, simply type, say, "cubic(3)".

B.4.2 *M-File Function*

When the function becomes too complex to be specified in the main code, you can write an independent M-file in the format of ".m" to create these functions. For instance:
FUNCTION $[y] =$ objective(x, c)
$y = 1 + exp(-c * x) - log(x);$

END
In the main code, call this function by typing, say, “objective(3, 1)". This is particularly helpful when you need to estimate some parameters that involves potentially minimizing complex objective functions to maximize a goodness of fit. An example of such a problem is in Chapter 3.

B.5 Graphing

Graphing is the subject of Chapter 5.

B.5.1 Plot

In MATLAB, two-dimensional line graphs can be created by the command PLOT. For instance, to plot a vector y against a vector x, simply use the command PLOT(x, y). Examples of time series plots are in 5 and all over the primer.
To add further features into the figure, consider graphing for two functions:
$F(x) = log(x)$ and $g(x) = 1 + e^{-cx}$.

```
% Create the Grid for Graphing
x_lb = 2; % lower bound of the grid
x_ub = 4; % upper bound of the grid
nx = 101; % the number of points in the grid for x

% create a grid for x on [x_lb, x_ub] with the number of points: nx
x_grid = linspace(x_lb, x_ub, nx);
c = 1.0; % parameter controlling the curvature of: 1 + exp(- c x)

% Graphing: Separately For Each Individual Functions
figure(1); % creates a new figure window
plot( x_grid, log(x_grid), x_grid, 1+ exp(- c * x_grid) ); % plot the two functions
title('Graphing For Each Individual Functions'); % make a title for this figure
xlabel('x'); % label for x axis
ylabel('log(x) and 1+ exp(-c*x)'); % label for y axis
legend('log(x)', '1+ exp(-c*x)'); % legends for each function
```

The default fonts on MATLAB graphs are small. These can be changed using the FONTSIZE command. So, in the above example to change the font size for the title to 18 just write:

title('Graphing For Each Individual Functions', 'FontSize', 18);

One can do the same thing for the axis labels.

More detailed graphing features of the command plot can be found in the M-file:
Ex_graphing.m.

B.5.2 *Tool Kit of Graphing Commands*

In addition, MATLAB offers a large choice set of graphing options – beyond PLOT–to use. Some handy tools are listed as follows:
BAR: create a bar chart. An example of a bar chart is in Chapter 5.
HIST: create a histogram. Chapter 9 provides an example of this.
PIE: create a pie chart. Chapter 5 also contains an example of a pie chart
SURF: three-dimension graphing. This is done in combination with the MESHGRID command that creates a mesh on the (x, y) plane. This command was used in Chapter 8 to generate a 3D quadratic approximation to a utility function.
WATERFALL: the waterfall plot.
PLOTYY: plot with multiple vertical axes.

B.6 *Solving Nonlinear Equations*

Nonlinear equations are covered in Chapter 2. The objective is to find the solution x^* which solves a nonlinear equation: $F(x^*) = 0$. Two scenarios and two methods are discussed here: smooth objective function solved by the Newton's method and nonsmooth objective function solved by the bisection method.

B.6.1 FSOLVE *and* FZERO

In MATLAB there is a straightforward built-in function to solve nonlinear equations: "FSOLVE". This section outlines the algorithm underlying this function and illustrates how this function is implemented.

FSOLVE: Algorithm

To be brief, the function FSOLVE is based on the Newton's method, i.e., it starts from an initial guess x_n, and updates the guess x_{n+1} recursively by the following rule:

$$F(x_n) + F_1(x_n)(x_{n+1} - x_n) = 0$$
$$\Longrightarrow x_{n+1} = x_n - [F_1(x_n)]^{-1}F(x_n)$$

In general FSOLVE works well when the function $F(\cdot)$ is smooth. In addition, as you can tell from the iteration rule above, the initial guess is critical in governing the computation efficiency. The solution of FSOLVE can be of any (finite) dimension. For the scenario of single-variable nonlinear equation, an alternative is the command FZERO, with analogous procedure to implement as FSOLVE. FZERO can only solve one equation in one unknown.

FSOLVE: Implementation

The general form to implement FSOLVE is:

x_star = FSOLVE(@(x) objective(x, c), x_initial)

There are five ingredients to implement FSOLVE:
objective(x, c): the nonlinear equation $F(x; c) = 0$ under consideration.
c: the exogenous parameter in this nonlinear equation.
x: the endogenous unknown to obtain.
x_star: the solution to this nonlinear equation, i.e., $F(x_{\text{star}}; c) = 0$.
x_initial: our initial guess for the solution x_star.

Of course, a variety of options can be added to the function FSOLVE. For example, to control for the criterion of convergence, use the command OPTIMSET:
opt = OPTIMSET('Tolfun', 1e-8)
x_star = FSOLVE(@(x) objective(x, c), x_initial, opt)

As a cookbook illustration, consider the following nonlinear equation:

$$\ln(x) = 1 + e^{-cx}$$

The procedure to solve for this equation is delineated in the M-file: **Ex_fsolve.m**.

B.6.2 *Bisection*

Unfortunately, Newton's method does not perform well when the objective function $F(\cdot)$ is not smooth, and a potential alternative solution is the bisection method.

The bisection method applies when our objective function $F(\cdot)$ is defined on an interval $[\underline{x}, \bar{x}]$ where $F(\underline{x}) \cdot F(\bar{x}) < 0$, i.e., the value of the objective function has opposite signs at the two boundaries of the interval in which it is defined. To apply the bisection method, we start from the initial guesses for the minimum and maximum values of the solution, and update our guess in accordance with the sign of the slope of the objective function.

To demonstrate how the bisection works, let's consider the same nonlinear equation:

$$\ln(x) = 1 + e^{-cx}$$

A detailed implementation can be found in the M-file: **Ex_bisection.m**. The FZERO command is based (partially) on the bisection method and is used to solve the monopoly problem in Section 2.7 of Chapter 2.

B.7 *Minimization (or Maximization)*

The maximization and minimization of functions is detailed in Chapter 3. Suppose that a solution is sought to a problem of the following form

$$\min_{\underline{x} \leq x \leq \overline{x}} \{F(x)\}.$$

Here x is constrained to lie in the interval $[\underline{x}, \overline{x}]$, where $\underline{x}$ and $\overline{x}$ are the lower and upper bounds on the minimization problem. The function can be minimized in MATLAB by calling the function FMINBND. The syntax to use fminbnd is

x = FMINBND(@function, lower bound, upper bound),

where x is the solution from the minimization routine. Suppose that one wants to maximize $F(x)$ on the domain $[\underline{x}, \overline{x}]$. This maximization problem can be transformed into a minimization problem as follows

$$\max_{\underline{x} \leq x \leq \overline{x}} \{F(x)\} = \min_{\underline{x} \leq x \leq \overline{x}} \{-F(x)\}.$$

So, one would just need to put a negative sign in front of the objective function.

The call to the particle swarm algorithm is similar. Now, one would write

x = PARTICLESWARM(@function, 1, lower bound, upper bound),

where the 1is telling the algorithm that there is one choice variable; i.e., the algorithm can be also be used when there is many choice variables and here lower bound and upper bound would be vectors of numbers.

Minimization routines are often used to maximize the goodness of fit for model; i.e., that is minimize the model's prediction errors. Examples are provided in Chapters 3 and 4.

B.8 *Roots of Polynomials*

Suppose that one wants to find the roots of an nth-order polynomial. This involves finding the n values of x that solve the equation.

$$a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x^1 + a_0 = 0.$$

This can be done in MATLAB using the ROOTS command. To use this just specify the vector $a = [a_n, a_{n-1}, \cdots, a_0]$, where the coefficients are in descending order and type the command ROOTS(a). The roots of polynomials are often computed to obtain the solution to linear difference equations. This is done in Chapter 8 to compute the dynamics for the neoclassical growth model.

B.9 Numerical Integration

Suppose that you have a function $F(x)$ that you want to integrate over the range $[a, b]$; i.e., you desire to computer $\int_a^b F(x)dx$. This can be done by using the command INTEGRAL(F, a, b). Chapter 8 discusses numerical integration.

B.10 Eigenvector decomposition

To find the eigenvalues and their corresponding eigenvectors of a square matrix A, the command $[e, \varepsilon]$ =EIG(A) returns the $n \times n$ diagonal matrix

$$\varepsilon = \begin{bmatrix} \varepsilon_1 & & 0 \\ & \ddots & \\ 0 & & \varepsilon_n \end{bmatrix}$$

of eigenvalues ε_i and the $n \times n$ matrix

$$e = \begin{bmatrix} e_1 & \dots & e_n \end{bmatrix}$$

whose $n \times 1$ columns are the corresponding right eigenvectors, so that $Ae = e\varepsilon$. Eigenvalues can be used to calculate the stationary distribution of a Markov chain. See Chapter 8 for more detail.

B.11 Polynomial Curve Fitting

A polynomial can be fit to data using the POLYFIT command. Given a set of data points (x, y) the coefficients, p, for a n-order polynomial can be obtained by writing POLYFIT(x, y, n). The polynomial has the form

$$p(x) = p_1 x^n + \cdots + p_n x + p_{n+1}.$$

Note MATLAB returns the coefficients in a decending order starting with the term on x^n. See Chapter 8 for more on a curve fitting polynomial.

B.12 Interpolation

Interpolation is covered in Chapter 8. MATLAB offers interpolation routines to add new data points within a range of a set of known data points for one to N-dimensional gridded data through commands INTERP1, INTERP2, INTERP3, or INTERPN. The variable at the end of INTERP refers to the number of arguments the function has. Focus on the two-dimensional case, where $y = f(x_1, x_2)$ is the function evaluated at each sample point (x_1, x_2) and y, x_1, and x_2 are $n \times n$ matrices. The command y^{query} =INTERP2$(x_1, x_2, y, x_1^{query}, x_2^{query}, method)$ returns interpolated values of y at the query points $(x_1^{query}, x_2^{query})$, using the interpolation method defined by $method =' linear',' nearest',' cubic',' makima', spline'$. Here the variables y^{query}, x_1^{query}, and x_2^{query} are vectors. Various interpolation schemes can be used. 'linear' refers so piecewise linear interpolation and 'spline' denotes standard cubic spline interpolation. 'Cubic' and 'makima' are variants of cubic spline interpolation. 'Nearest' is a cousin of piecewise linear interpolation. See Chapter 8 for a discussion on interpolation.

B.13 Random Number Generation

To call up a $1 \times n$ vector normal random numbers with mean mu and standard deviation sigma just type in NORMRND(mu, sigma,1,n). To seed the random number generation use the statement RNG=int, where int is some positive integer, just before you call NORMRND. If you don't do this, your random numbers will change every time you run your program. You can also call up uniformly distributed random numbers. RAND(n,m) will yield a $n \times m$ matrix of uniformly distributed random numbers on the interval $[0, 1]$. RANDI(a,b,n) will return n integers that are uniformly distributed on the interval $[a, b]$. Again, you should seed these random number generators. Examples of the use of random number generation are the Slutsky's business cycle model and the Lucas welfare cost of business cycle calculation discussed in Chapter 8.

B.14 Complex Numbers

Complex numbers are easy in MATLAB. A complex number has the form $a + bi$, where a and b are real numbers and i is the imaginary part or where $i = \sqrt{-1}$. In MATLAB a complex number is written in exactly this way, where I is reserved for $\sqrt{-1}$. To recover the coefficent b on a complex number $x = a + bi$, just type IMAG(x).

B.15 Descriptive Statistics

Suppose one has two vectors of equal size, x and y, containing data. To compute the standard deviation of x use the command STD(x). To compute the correlation between x and y just write CORRCOEF(x,y). To plot a histogram for x use the command HISTOGRAM(x). This command can be used in conjunction with TITLE, XLABEL, and YLABEL to generate a title and axis labels for the histogram. Such commands are used in Chapter 9.

B.16 Symbolic Toolbox

To declare variable x symbolically use the command SYMS x. Differentiating the function $F(x)$ with respect to x is easy, just write DIFF(F, x). The equation $eqn(x) = 0$ can be solved for x by commanding SOLVE(eqn,x). Last, a solution S can be simplified using the line SIMPLIFY(S).

C Introduction to Julia

Julia is a high-level programing language well suited for numerical analysis. One of its main advantages resides in its speed as Julia compiles all code to machine code before running it. This brief introduction highlights the different steps needed to get us up to speed with the code provided in the book.

C.1 Choosing an IDE

The first step before using Julia is to choose an Integrated Development Environment (IDE), a platform to help you write and run code efficiently. There are free options available, the most popular ones at the moment are Juno for Atom and Julia for VS Code. These editors offer an interface that is similar to Matlab, including a Workspace displaying packages in use, structures and variables created, an Editor to write code, and Julia's command line REPL (standing for Read, Execute, Print, Loop), where the code's output is printed.

C.2 Installing and Using Packages

Julia has a built-in package manager that allows the user to install packages and to call them when needed to run code. Packages are installed by typing "Alt]" in the REPL and then "add PACKAGE_NAME" and enter. Alternatively, you can type "using Pkg" and then enter to call the package manager, followed by "Pkg.add("PACKAGE_NAME")" and enter.

Key packages

For data analysis: DataFrames.jl offers tools for working with tabular data. CSV.jl is a package for reading csv data.

For differentiation: ForwardDiff.jl uses forward mode automatic differentiation to evaluate the derivatives of functions and compute gradients, Jacobian and Hessian matrices. FiniteDifferences.jl offers an alternative method to estimate derivatives with finite differences.

For solving nonlinear problems: ROOTS.JL offers algorithms to find the roots of continuous scalar funciton of a single real variable (e.g. using Bisection method, Brent's method and derivative-free methods). NLSOLVE.JL provides algorithms to solve systems of nonlinear equations, including mixed complementarity problems.

For solving optimization problems: OPTIM.JL offers algorithms to solve univariate and multivariate optimization problem with box constraints and includes methods such as simulated annealing and particle swarm. BLACKBOXOPTIM.JL provides global optimization algorithms that do not require the objective function to be differentiable. NLOPT.JL is an interface offering a suit of different optimization algorithms. JUMP.JL is another optimization interface for a number of open-source and commercial solvers targeted at constrainted problems.

For plotting: PLOTS.JL is a data visualization interface that can be combined with other backends such as PGFPLOTSX.JL, PLOTLYJS.JL, and PYPLOT.JL.

C.3 *Basic Commands*

Housekeeping

Prior to the main body of the coding work, there are two basic commands commonly used for housekeeping:
EXIT(): removes all packages, structures, and variables from the current workspace.
CLEARCONSOLE(): clears all input and output from the REPL.

Loading data

To import data

Stopwatch timer

To monitor the

Lines in a file

Julia does not require the use of three dots ("... ") continue statements across lines.

Table C.3.1: Matlab-Julia cheat-sheet

Matlab	Julia	Packages
Vectors and matrices		

`% Row vector`
`A = [1 2 3]`
`% Column vector`
`A = [1; 2; 3]`
`% Matrix`
`A = [1 2; 3 4]`
`% Matrix of zeros`
`A = zeros(2, 2)`
`% Matrix of ones`
`A = ones(2, 2)`
`% Identity matrix`
`A = eye(2,2)`
`% Diagonal matrix`
`A = diag([1 2 3])`
`% Linearly spaced vector`
`A = linspace(x_ini, x_final, n)`
	`% Row vector`
	`A = [1 2 3]`
	`% Column vector`
	`A = [1 2 3]'`
	`% Matrix`
	`A = [1 2; 3 4]`
	`% Matrix of zeros`
	`A = zeros(2, 2)`
	`% Matrix of ones`
	`A = ones(2, 2)`
	`% Identity matrix`
	`A = I`
	`% Diagonal matrix`
	`A = Diagonal([1, 2, 3])`
	`% Linearly spaced vector`
	`A = range(x_ini, x_final, length = n)`	

Interpolations		
Distributions		

References

S. Rao Aiyagari. Uninsured idiosyncratic risk and aggregate savings. *Quarterly Journal of Economics*, 109(3):659–684, 1994.

Richard Bellman. *Dynamic Programming*. Princeton University Press, Princeton, NJ, 1957.

Truman Bewley. The permanent income hypothesis: A theoretical formulation. *Journal of Economic Theory*, 16(2):252–292, 1977.

Truman Bewley. *The Optimum Quantity of Money*. *In:* Kareken, John, Wallace, Neil (Eds.), Models of Monetary Economics. Federal Reserve Bank of Minneapolis, Minnesota, 1980.

Timo Boppart, Per Krusell, and Kurt Mitman. Exploiting mit shocks in heterogeneous-agent economies: The impulse response as a numerical derivative. *Journal of Economic Dynamics & Control*, 89:68–92, 2018.

William Brock and Leonard Mirman. Optimal economic growth and uncertainty: The discounted case. *Journal of Economic Theory*, 4(3): 479–513, 1972.

Victor Bryant. *Metric Spaces: Iteration and Application*. Cambridge University Press, Cambridge, U.K., 1985.

Christopher D. Carroll. The method of endogenous gridpoints for solving dynamic stochastic optimization problems. *Economics Letters*, 3 (91):312–320, 2006.

Alpha C. Chiang. *Fundamental Methods of Mathematical Economics*. McGraw-Hill Book Company, New York, NY, 3 edition, 2011.

Wilbur John Coleman, II. Equilibrium in a production economy with an income tax. *Econometrica*, 59(4):1091–1104, 1991.

Morris H. DeGroot. *Probability and Statistics*. Addison-Wesley Publishing Company, Reading, MA, 1975.

Wouter J. den Haan and Albert. Marcet. Solving the stochastic growth model by parameterizing expectations. *Journal of Business and Economic Statistics*, 8(1):31–34, 1990.

Ray C. Fair and John B. Taylor. Solution and maximum likelihood estimation of dynamic nonlinear rational expectations models. *Econometrica*, 51(4):1169–1185, 1983.

Daniel Feenberg and Elisabeth Coutts. An introduction to the taxsim model. *Journal of Policy Analysis and Management*, 12(1):189–194, 1993.

Jeremy Greenwood and Karen A. Kopecky. Measuring the welfare gain from personal computers. *Economic Inquiry*, 51(1):336–347, 2013.

Jeremy Greenwood, Zvi Hercowitz, and Gregory W. Huffman. Investment, capacity utilization, and the real business cycle. *American Economic Review*, 78(3):402–417, 1988.

Jeremy Greenwood, Nezih Guner, and Karen A. Kopecky. The wife's protector: A quantitative theory linking contraceptive technology with the decline in marriage. *Handbook of Historical Economics*, 2021a.

Jeremy Greenwood, Nezih Guner, and Ricardo Marto. The great transition: Kuznets facts for family economists. *Handbook of Historical Economics*, 2021b.

Gary D. Hansen and Edward C. Prescott. Malthus to solow. *American Economic Review*, 92(4):1205–1217, 2002.

John R. Hicks. *Value and Capital*. Oxford University Press, Oxford, 1939.

John R. Hicks. The rehabilitation of consumers' surplus. *Review of Economic Studies*, 8(2):108–116, 1941.

Robert J. Hodrick and Edward C. Prescott. Postwar u.s. business cycles: An empirical investigation. *Journal of Money, Credit and Banking*, 29 (1):1–16, 1997.

Mark Hugget. The risk-free rate in heterogeneous-agent incomplete-insurance economies. *Journal of Economic Dynamics & Control*, 17(5-6): 953–969, 1993.

Mark Hugget. The one sector growth model with idiosyncratic shocks: Steady states and dynamics. *Journal of Monetary Economics*, 39(3): 385–403, 1997.

Ayşe Imrohoroğlu. Cost of business cycles with indivisibilities and liquidity constraints. *Journal of Political Economy*, 97(6):1364–1383, 1989.

Kenneth L. Judd, Lilia Maliar, and Serguei Maliar. Numerically stable and accurate stochastic simulation approaches for solving dynamic economic models. *Quantitative Economics*, 2(2):173–210, 2011.

Tjalling C. Koopmans. Measurement without theory. *Review of Economics and Statistics*, 29(3):161–172, 1947.

Finn E. Kydland and Edward C. Prescott. Time to build and aggregate fluctuations. *Econometrica*, 50(6):1345–1370, 1982.

John Laitner. Random earnings differences, lifetime liquidity constraints, and altruistic intergenerational transfers. *Journal of Economic Theory*, 58(2):135–170, 1992.

Wassily W. Leontief. Input-output economics. *Scientific American*, 185 (4):15–21, 1951.

Robert E. Lucas, Jr. Models of business cycles. 1987.

Robert E. Lucas, Jr. and Edward C Prescott. Investment under uncertainty. *Econometrica*, 5(39):659–681, 1971.

Rajnish Mehra and Edward C. Prescott. The equity premium: A puzzle. *Journal of Monetary Economics*, 15(2):145–161, 1985.

Edward C. Prescott. Why do americans work so much more than europeans? *Quarterly Review*, 28(1):2–13, 2004.

Edward C. Prescott and Graham V. Chandler. *Calibration*. The New Palgrave: A Dictionary of Economics. Palgrave Macmillan, 2 edition, 2008.

Frank P. Ramsey. A mathematical theory of saving. *Economic Journal*, 38(152):543–559, 1928.

K. Geert Rouwenhorst. *Asset Pricing Implications of Equilibrium Business Cycle Models*. Thomas F. Cooley, Princeton, NJ, Princeton University Press, 1995.

Jonathan A. Schwabish. An economist's guide to visualizing data. *Journal of Economic Perspectives*, 28(Winter):209–34, 2014.

Eugen Slutsky. Sulla teoria del bilancio del consumatore. *Giornale degli Economisti e Rivista di Statistica*, 51(1):1–26, 1915.

Eugen Slutsky. The summation of random causes as a source of cyclic processes. *Econometrica*, 5(2):105–146, 1937.

Nancy L. Stokey and Robert E. Lucas, Jr. with Edward C. Prescott. *Recursive Methods in Economic Dynamics*. Harvard University Press, Cambridge, MA, 1986.

Richard Stone and Giovanna. Stone. *National Income and Expenditure*. Bowes and Bowes, London, 1966.

Edward R. Tufte. *The Visual Display of Quantitative Information*. Graphics Press, Cheshire, Connecticut, 2001.

Edmund T. Whittaker. On a new method of graduation. *Proceedings of the Edinburgh Mathematical Society*, 41(February):63–73, 1923.