Macroeconomic Theory and Policy

Preliminary Draft

David Andolfatto
Simon Fraser University
dandolfa@sfu.ca

Contents

Preface

The field of macroeconomic theory has evolved rapidly over the last quarter century. A quick glance at the discipline's leading journals reveals that virtually the entire academic profession has turned to interpreting macroeconomic data with models that are based on *microeconomic foundations*. Unfortunately, these models often require a relatively high degree of mathematical sophistication, leaving them largely inaccessible to the interested lay person (students, newspaper columnists, business economists, and policy makers). For this reason, most public commentary continues to be cast in terms of a language that is based on simpler 'old generation' models learned by policymakers in undergraduate classes attended long ago.

To this day, most introductory and intermediate textbooks on macroeconomic theory continue to employ old generation models in expositing ideas. Many of these textbooks are written by leading academics who would not be caught dead using any of these models in their research. This discrepancy can be explained, I think, by a widespread belief among academics that their 'new generation' models are simply too complicated for the average undergraduate. The use of these older models is further justified by the fact that they do in some cases possess hidden microfoundations, but that revealing these microfoundations is more likely to confuse rather than enlighten. Finally, it could be argued that one virtue of teaching the older models is that it allows students to better understand the language of contemporary policy discussion (undertaken by an old generation of former students who were taught to converse in the language of these older models).

While I can appreciate such arguments, I do not in general agree with them. It is true that the models employed in leading research journals are complicated. But much of the *basic intuition* embedded in these models can often be exposited with simple diagrams (budget sets and indifference curves). The tools required for such analysis do not extend beyond what is regularly taught in a good undergraduate microeconomics course. And while it is true that many of the older generation models possess hidden microfoundations, I think that it is mistake to hide these foundations from students. Among other things, a good understanding of a model's microfoundations lays bare its otherwise hidden assumptions, which is useful since it renders clearer the model's limitations and forces the student to think more carefully. A qualified professional can get away with using 'short cut' models with hidden microfoundations, but in the hands of a layman, such models can be the source of much mischief (bad policy advice). I am somewhat more sympathetic to the last argument concerning language. A potential pitfall of teaching macroeconomics using a modern language is that students may be left in a position that leaves them unable to decipher the older language still widely employed in policy debates. Here, I think it is up to the instructor to draw out the mapping between old and new language whenever it might be useful to do so. Unfortunately, translation is time-consuming. But it is arguably a necessary cost to bear, at least, until the day the old technology is no longer widely in use.

To understand why the new generation models constitute a better technology, one needs to understand the basic difference between the two methodological approaches. The old generation models rely primarily on *assumed* behavioral relationships that are simple to analyze and seem to fit the historical data reasonably well. No formal explanation is offered as to *why* people might rationally choose follow these rules. The limitations of this approach are twofold. First, the assumed behavioral relations (which can fit the historical data well) often seemed to 'break down' when applied to the task of predicting the consequences of new government policies. Second, the behavioral relations do not in themselves suggest any natural criterion by which to judge whether any given policy makes people better or worse off. To circumvent this latter problem, various *ad hoc* welfare criteria emerged throughout the literature; e.g., more unemployment is bad, more GDP is good, a current account deficit is bad, business cycles are bad, and so on. While all of these statements sound intuitively plausible, they constitute little more than bald assertions.

In contrast, the new generation of models rely more on the tools of microeconomic theory (including game theory). This approach assumes that economic decisions are made for a reason. People are assumed to have a well-defined objective in life (represented by preferences). Various constraints (imposed by nature, markets, the government, etc.) place restrictions on how this objective can be achieved. By assuming that people try to do the best they can subject to these constraints, optimal behavioral rules can be *derived* instead of assumed. Macroeconomic variables can then be computed by summing up the actions of all individuals. This approach has at least two main benefits. First, to the extent that the deep parameters describing preferences and constraints are approximated reasonably well, the theory can provide reliable predictions over any number of hypothetical policy experiments. Second, since preferences are modeled explicitly, one can easily evaluate how different policies may affect the welfare of individuals (although, the problem of constructing a social welfare function remains as always). As it turns out, more unemployment is not always bad, more GDP is not always good, a current account deficit is not always bad, and business cycles are not necessarily bad either. While these results may sound surprising to those who are used to thinking in terms of old generation models, they emerge as logical outcomes with intuitive explanations when viewed from the perspective of modern macroeconomic theory.

The goal of this textbook is to provide students with an introduction to the microfoundations of macroeconomic theory. As such, it does not constitute a survey of all the different models that inhabit the world of modern macroeconomic research. It is intended primarily as an exposition designed to illustrate the *basic idea* that underlies the modern research methodology. It also serves to demonstrate how the methodology can be applied to interpreting macroeconomic data, as well as how the approach is useful for evaluating the economic and welfare consequences of different government policies. The text is aimed at a level that should be accessible to any motivated third-year student. A good understanding of the text should pay reasonable dividends, especially for those who are inclined to pursue higher-level courses or possibly graduate school. But even for those who are not so inclined, I hope that the text will at least serve as interesting food for thought.

Of course, this is not the first attempt to bring the microeconomic foundations of macroeconomic theory to an undergraduate textbook. An early attempt is to be found in: *Macroeconomics: A Neoclassical Introduction*, by Merton Miller and Charles Upton (Richard D. Irwin, Inc.,1974). This is still an excellent text, although it is by now somewhat dated. More recent attempts include: *Macroeconomics,* by Robert Barro (John Wiley and Sons, Inc., 1984); *Macroeconomics: An Integrated Approach*, by Alan Auerbach and Lawrence Kotlikoff (MIT Press, 1998); and *Macroeconomics*, by Stephen Williamson (Addison Wesley, 2002).

These are all excellent books written by some of the profession's leading academics. But like any textbook, they each have their particular strengths and weaknesses (just try writing one yourself). Without dwelling on the weaknesses of my own text, let me instead highlight what I think are its strengths. First, I present the underlying choice problems facing individuals explicitly and systematically throughout the text. This is important, I think, because it serves to remind the student that to understand individual (and aggregate) behavior, one needs to be clear about what motivates and constrains individual decision-making. Second, I present simple mathematical characterizations of optimal decision-making and equilibrium outcomes, some of which can be solved for analytically with high-school algebra. Third, I try (in so far that it is possible) to represent optimal choices and equilibrium outcomes in terms of indifference curve and budget set diagrams. The latter feature is important because the position of an indifference curve can be used to assess the welfare impact of various changes in the economic or physical environment. Fourth, through the use of examples and exercises, I try to show how the theory can be used to interpret data and evaluate policy.

The text also contains chapters that are not commonly found in most textbooks. Chapter 7, for example, the modern approach to labor market analysis, which emphasizes the gross flows of workers across various labor market states and interprets the phenomenon of unemployment as an equilibrium outcome.

Chapter 10 develops a simple, but explicit model of fiat money and Chapter 11 utilizes this tool to discuss nominal exchange rates (emphasizing the problem of indeterminacy). Finally, the section on economic development extends beyond most texts in that it includes: a survey of technological developments since classical antiquity; presents the Malthusian model of growth; introduces the concept of endogenous fertility choice; and addresses the issue of special interests in the theory of productivity differentials (along with the usual topics, including the Solow model and endogenous growth theory).

I realize that it may not be possible to cover every chapter in a semester long course. I view Chapters 1-6 as constituting 'core' material. Following the exposition of this material, the instructor may wish to pick and choose among the remaining chapters depending on available time and personal taste.

At this stage, I would like to thank all my past students who had to suffer through preliminary versions of these notes. Their sharp comments (and in some cases, biting criticisms) have contributed to a much improved text. I would especially like to thank Sultan Orazbayez and Dana Delorme, both of whom have spent hours documenting and correcting the typographical errors in an earlier draft. Thoughtful comments were also received from Bob Delorme and Janet Hua. I am also grateful for the thoughtful suggestions offered by several anonymous reviewers. This text is still very much a work in progress and I remain open to further comments and suggestions for improvement. If you are so inclined, please send them to me via my email address: `dandolfa@sfu.ca`

Part I

Macroeconomic Theory: Basics

Chapter 1

The Gross Domestic Product

1.1 Introduction

The GDP measures the value of an economy's production of goods and services (*output*, for short) over some interval of time. A related statistic, called the *per capita GDP*, measures the value of production *per person*. Economists and policymakers care about the GDP (and the per capita GDP in particular) because material living standards depend largely on what an economy produces in the way of final goods and services. Residents of an economy that produces more food, more clothes, more shelter, more machinery, etc., are likely to be better off (at least, in a material sense) than citizens belonging to some other economy producing fewer of these objects. As we shall see later on, the link between an economy's per capita GDP and individual well-being (welfare) is not always exact. But it does seem sensible to suppose that by and large, higher levels of production (per capita) in most circumstances translate into higher material living standards.

Definition: The GDP measures the value of all final goods and services (output) produced domestically over some given interval of time.

Let us examine this definition. First of all, note that the GDP measures only the production of *final* goods and services; in particular, it does not include the production of *intermediate* goods and services. Loosely speaking, intermediate goods and services constitute materials that are used as inputs in the construction final goods or services. Since the market value of the final output already reflects the value of its intermediate products, adding the value of intermediate materials to the value of final output would overstate the true value of production in an economy (one would, in effect, be double counting). For example, suppose that a loaf of bread (a final good) is produced with flour (an intermediate good). It would not make sense to add the value of flour separately in the calculation of GDP since the flour has been 'consumed' in process of making bread and since the market price of bread already reflects the value of the flour that was used in its production.

Now, consider the term 'gross' in the definition of GDP. Economists make a distinction between the gross domestic product and the *net* domestic product (NDP). The NDP essentially corrects the GDP by subtracting off the value of the capital that depreciates in the process of production. Capital depreciation is sometimes also referred to as capital consumption.

Definition: The NDP is defined as the GDP less capital consumption.

A case could be made that the NDP better reflects an economy's level of production since it takes into account the value of capital that is consumed in the production process. Suppose, for example, that you own a home that generates $12,000 of rental income (output in the form of shelter services). Imagine further that your tenants are university students who (over the course of several parties) cause $10,000 in damage (capital consumption). While your gross income is $12,000 (a part of the GDP), your income net of capital depreciation is only $2,000 (a part of the NDP). If you are like most people, you probably care more about the NDP than the GDP. In fact, environmental groups often advocate the use of an NDP measure that defines capital consumption broadly to include 'environmental degradation.' Conceptually, this argument makes sense, although measuring the value of environmental degradation can be difficult in practice.

Finally, consider the term 'domestic' in the definition of GDP. The term 'domestic' refers to the economy that consists of all production units (people and capital) that reside within the national borders of a country. This is not the only way to define an economy. One could alternatively define an economy as consisting of all production units that belong to a country (whether or not these production units reside in the country or not). For an economy defined in this way, the value of production is called the *Gross National Product* (GNP).

Definition: The GNP measures the value of all final goods and services (output) produced by citizens (and their capital) over some given interval of time.

The discrepancy between GDP and GNP varies from country to country. In Canada, for example, GDP has recently been larger than GNP by only two or three percent. The fact that GDP exceeds GNP in Canada means that the value of output produced by foreign production units residing in Canada is larger than the value of output produced by Canadian production units residing outside of Canada. While the discrepancy between GDP and GNP is relatively small for Canada, the difference for some countries can be considerably larger.

1.2 How GDP is Calculated

Statistical agencies typically estimate an economy's GDP in two ways: the *income approach* and the *expenditure approach*.[1] In the absence of any measurement errors, both approaches will deliver exactly the same result. Each approach is simply constitutes a different way of looking at the same thing.

1.2.1 The Income Approach

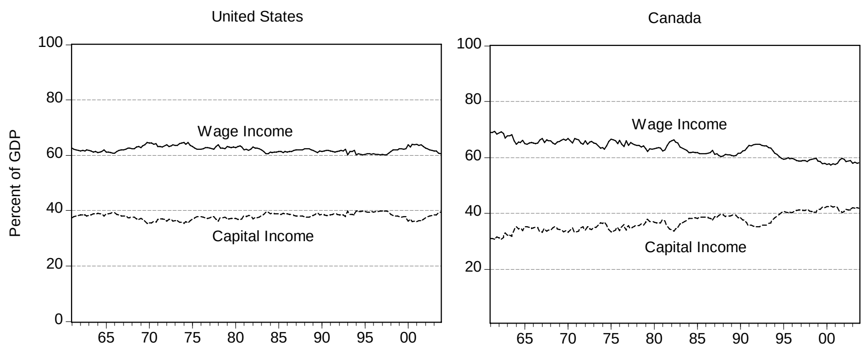
As the name suggests, the income approach calculates the GDP by summing up the income earned by domestic factors of production. Factors of production can be divided into two broad categories: capital and labor. Let R denote the income generated by capital and let L denote the income generated by labor. Then the *gross domestic income* (GDI) is defined as:

$$GDI \equiv L + R.$$

Figure 1.1 plots the ratio of wage income as a ratio of GDP for the United States and Canada over the period 1961-2002. From this figure, we see that wage income constitutes approximately 60% of total income, with the remainder being allocated to capital (broadly defined). Note that for these economies, these ratios have remained relatively constant over time (although there appears to be a slight secular trend in the Canadian data over this sample period). One should keep in mind that the distribution of income across factors of production is not the same thing as the distribution of income across individuals. The reason for this is that in many (if not most) individuals own at least some capital (either directly, through ownership of homes, land, stock, and corporate debt, or indirectly through company pension plans).

[1] There is also a third way, called the *value-added* or *product approach*, that I will not discuss here.

FIGURE 1.1
GDP Income Components
United States and Canada
1961.1 - 2003.4



1.2.2 The Expenditure Approach

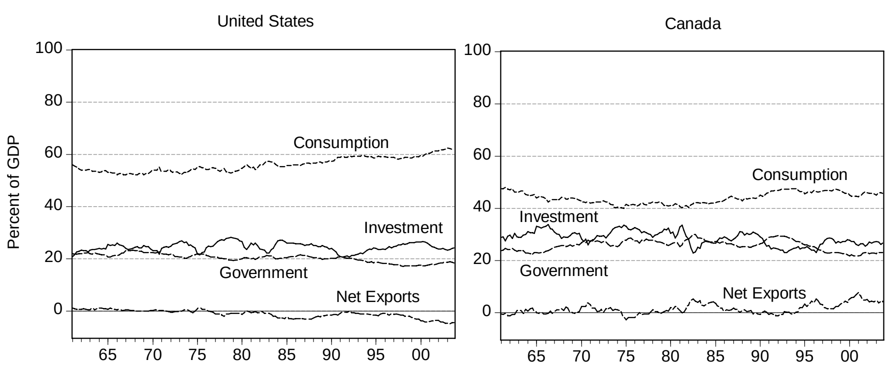
In contrast to the income approach, the expenditure approach focuses on the uses of GDP across various expenditure categories. Traditionally, these expenditure categories are constructed by dividing the economy into four sectors: (1) a household sector; (2) a business sector; (3) a government sector; and (4) a foreign sector. Categories (1) and (2) can be combined to form the *private* sector. The private sector and the government sector together form the *domestic* sector.

Let C denote the expenditures of the household sector on consumer goods and services (consumption), including imports. Let I denote the expenditures of the business sector on new capital goods and services (investment), including imports. Let G denote the expenditures by the government sector on goods and services (government purchases), including imports. Finally, let X denote the expenditures on domestic goods and services undertaken by residents of the foreign sector (exports). Total expenditures are thus given by $C+I+G+X$. Of course, some of the expenditures on C, I and G consist of spending on imports, which are obviously not goods and services that are produced domestically. In order to compute the gross *domestic* expenditure (GDE), on must subtract off the value of imports, M. If one defines the term $NX \equiv X - M$ (net exports), then the GDE is given by:

$$GDE \equiv C + I + G + NX.$$

Figure 1.2 plots the expenditure components of GDP (as a ratio of GDP) for the United States and Canada over the period 1961-2002. Once again, it is interesting to note the relative stability of these ratios over long periods of time. To a first approximation, it appears that private consumption expenditures (on services and nondurables) constitute between 50–60% of GDP, private investment expenditures constitute between 20–30% of GDP, government purchases constitute between 20–25% of GDP, with NX averaging close to 0% over long periods of time. Note, however, that in recent years, the United States has been running a negative trade balance while Canada has been running a positive trade balance.

FIGURE 1.2
GDP Expenditure Components
United States and Canada
1961.1 - 2003.4



1.2.3 The Income-Expenditure Identity

So far, we have established that $GDP \equiv GDI$ and $GDP \equiv GDE$. From these two equivalence relations, it follows that $GDE \equiv GDI$. In other words, aggregate expenditure is equivalent to aggregate income, each of which are equivalent to the value of aggregate production. One way to understand why this must be true is as follows. First, any output that is produced must also be purchased (additions to inventory are treated purchases of new capital goods, or investment spending). Hence the value of production must (by definition) be equal to the value of spending. Second, since spending by one individual constitutes income for someone else, total spending must (by definition) be equal to total income.

The identity $GDI \equiv GDE$ is sometimes referred to as the *income-expenditure identity*. Letting Y denote the GDI, most introductory macroeconomic textbooks express the income-expenditure identity in the following way:

$$Y \equiv C + I + G + X - M.$$

Note that since the income-expenditure identity is an identity, it *always* holds true. However, it is very important to understand what this identity does and does not imply. A natural inclination is to suppose that since the identity is always true, one can use it to make *theoretical* or *predictive* statements. For example, the identity seems to suggest that an expansionary fiscal policy (an increase in G) must necessarily result in an increase in GDP (an increase in Y). In fact, *the income-expenditure identity implies no such thing.*

To understand why this is the case, what one must recognize is that an identity is *not a theory* about the way the world works. In particular, the income-expenditure identity is nothing more than a *description* of the world; i.e., it is simply categorizes GDP into to its expenditure components and then exploits the fact that total expenditure is *by construction* equivalent to total income. To make predictions or offer interpretations of the data, one must *necessarily* employ some type of theory. As we shall see later on, an increase in G *may* or *may not* lead to an increase in Y, depending on circumstances. But whether or not Y is predicted to rise or fall, the income-expenditure identity will always hold true.

1.3 What GDP Does Not Measure

Before moving on, it is important to keep in mind what GDP does not measure both in principle and in practice (i.e., things that should be counted as GDP in principle, but may not be in practice).

In principle, GDP is supposed to measure the value of output that is in some sense 'marketable' or 'exchangeable' (even if it is not actually marketed or exchanged). For example, if you spend 40 hours a week working in the market sector, your earnings measure the market value of the output you produce. However, there are 168 hours in a week. What are you producing with your remaining 126 hours? Some of this time may be spent producing marketable output that is not exchanged in a market. Some examples here include the time you spend doing housework, mowing the lawn, and repairing your car, etc. Assuming that you do not like to do any of these things, you could contract out these chores. If you did, what you pay for such services would be counted as part of the GDP. But whether you contract out such services or not, they clearly have value and this value *should* be counted as part of the GDP (even if it is not always done so in practice).

The great majority of peoples' time, however, appears to be employed in the production of 'nonmarketable' output. Nonmarketable output may be either in the form of consumption or investment. As a consumption good, a nonmarketable output is an object that is simultaneously produced and consumed by the individual producing it. An obvious example here is sleep (beyond what is necessary to maintain one's health). It is hard to get someone else to sleep for you. A wide variety of leisure activities fall into this category as well (imagine asking someone to go on vacation for you). As an investment good, a nonmarketable output is an object that remains physically associated with the individual producing it. Time spent in school accumulating 'human capital' falls into this category.[2] A less obvious example may also include time spent searching for work. Nonmarketable output is likely very large and obviously has value. However, it is not counted as part of an economy's GDP.

Another point to stress concerning GDP as a measure of 'performance' is that it tells us nothing about the distribution of output in an economy. At best, the (per capita) GDP can only give us some idea about the level of production accruing to an 'average' individual in the economy.

Finally, it should be pointed out that there may be a branch of an economy's production flow should be counted as GDP in principle, but for a variety of reasons, is not counted as such in practice. Ultimately, this problem stems with the lack of information available to statistical agencies concerning the production of marketable output that is either consumed by the producer or exchanged in 'underground' markets; see Appendix 1.A for details.

1.4 Nominal versus Real GDP

GDP was defined above as the value of output (income or expenditure). The definition did not, however, specify in which units 'value' is to be measured. In everyday life, the value of goods and services is usually stated in terms of *market prices* measured in units of the national currency (e.g., Canadian dollars). For example, the dozen bottles of beer you drank at last night's student social cost you \$36 (and possibly a hangover). The 30 hours you worked last week cost your employer \$300; and so on. If we add up incomes and expenditures in this manner, we arrive at a GDP figure measured in units of money; this measure is called the *nominal* GDP.

If market prices (including nominal exchange rates) remained constant over time, then the nominal GDP would make comparisons of GDP across time and countries an easy task (subject to the caveats outlined in Appendix 1.A). Unfortunately, as far as measurement issues are concerned, market prices do not remain constant over time. So why is this a problem?

The value of either income or expenditure is measured as the product of prices (measured in units of money) and quantities. It seems reasonable to suppose that material living standards are somehow related to quantities; and not the value of these quantities measured in money terms. In most economies

[2]Note that while the services of the human capital accumulated in this way may subsequently be rented out, the human capital itself remains embedded in the individual's brain. As of this writing, no technology exists that allows us to trade bits of our brain.

(with some notable exceptions), the general level of prices tends to grow over time; such a phenomenon is known as *inflation*. When inflation is a feature of the economic environment, the nominal GDP will rise even if the quantities of production remain unchanged over time. For example, consider an economy that produces nothing but bread and that year after year, bread production is equal to 100 loaves. Suppose that the price of bread ten years ago was equal to \$1.00 per loaf, so that the nominal GDP then was equal to \$100. Suppose further that the price of bread has risen by 10% per annum over the last ten years. The nominal GDP after ten years is then given by $(1.10)^{10}(\$100) = \$260$. Observe that while the nominal GDP is 2.6 times higher than it was ten years ago, the 'real' GDP (the stuff that people presumably care about) has remained constant over time.

Thus, while measuring value in units of money is convenient, it is also problematic as far as measuring material living standards. But if we can no longer rely on market prices denominated in money to give us a common unit of measurement, then how are we to measure the value of an economy's output? If an economy simply produced one type of good (as in our example above), then the answer is simple: Measure value in units of the good produced (e.g., 100 loaves of bread). In reality, however, economies typically produce a wide assortment of goods and services. It would make little sense to simply add up the level of individual quantities produced; for example, 100 loaves of bread, plus 3 tractors, and 12 haircuts does not add up to anything that we can make sense of.

So we return to the question of how to measure 'value.' As it turns out, there is no unique way to measure value. How one chooses to measure things depends on the type of 'ruler' one applies to the measurement. For example, consider the distance between New York and Paris. How does one measure distance? In the United States, long distances are measured in 'miles.' The distance between New York and Paris is 3635 miles. In France, long distances are measured in 'kilometers'. The distance between Paris and New York is 5851 kilometers. Thankfully, there is a fixed 'exchange rate' between kilometers and miles (1 mile is approximately 1.6 kilometers), so that both measures provide the same information. Just as importantly, there is a fixed exchange rate between miles *across* time (one mile ten years ago is the same as one mile today).

The phenomenon of inflation (or deflation) distorts the length of our measuring instrument (money) over time. Returning to our distance analogy, imagine that the government decides to increase the distance in a mile by 10% per year. While the distance between New York and Paris is currently 3635 miles, after ten years this distance will have grown to $(1.10)^{10}(3635) = 9451$ miles. Clearly, the increase in distance here is just an illusion (the 'real' distance has remained constant over time). Similarly, when there is an inflation, growth in the nominal GDP will give the illusion of rising living standards, even if 'real' living standards remain constant over time.

There are a number of different ways in which to deal with the measurement issues introduced by inflation. Here, I will simply describe one approach that is commonly adopted by statistical agencies. Consider an economy that produces n different goods and services. Let t denote the time-period (e.g., year) under consideration. Let x_t^i denote the quantity of good i produced at date t and let p_t^i denote the money price of good i produced at date t. Statistical agencies collect information on the expenditures made on each domestically produced good and service; i.e., $p_t^i x_t^i$, for $i = 1, 2, ..., n$ and for each year t. The gross domestic expenditure (measured in *current* dollars) is simply given by:

$$GDE_t = \sum_{i=1}^{n} p_t^i x_t^i.$$

Now, choose one year arbitrarily (e.g., $t = 1997$) and call this the *base year*. Then, the real GDP (RGDP) in any year t is calculated according to the following formula:

$$RGDP_t \equiv \sum_{i=1}^{n} p_{1997}^i x_t^i.$$

This measure is called the GDP (expenditure based) in terms of base year (1997) prices. In other words, the value of the GDP at date t is now measured in units of 1997 dollars (instead of current, or date t dollars). Note that by construction, $RGDP_{1997} = GDE_{1997}$.

As a by-product of this calculation, one can calculate the average level of prices (technically, the *GDP Deflator* or simply, the *price level*) P_t according to the formula:

$$P_t \equiv \frac{GDE_t}{RGDP_t}.$$

Note that the GDP deflator is simply an index number; i.e., it has no economic meaning (in particular, note that $P_{1997} = 1$ by construction). Nevertheless, the GDP deflator is useful for making comparisons in the price level across time. That is, even if $P_{1997} = 1$ and $P_{1998} = 1.10$ individually have no meaning, we can still compare these two numbers to make the statement that the price level rose by 10% between the years 1997 and 1998.

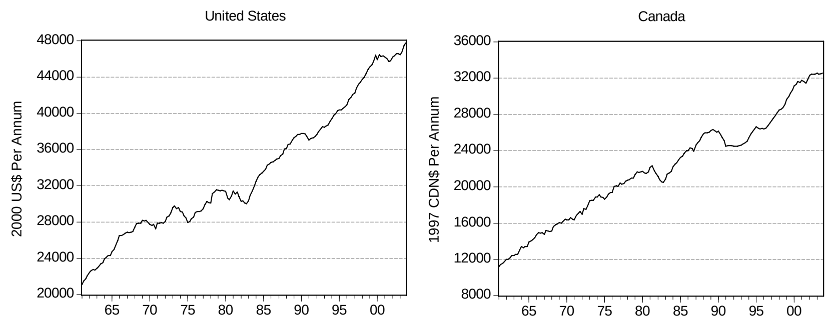
The methodology just described above is not fool-proof. In particular, the procedure of using base year prices to compute a measure of real GDP assumes that the structure of relative prices remains constant over time. To the extent that this is not true (it most certainly is not), then measures of the growth rate in real GDP can depend on the arbitrary choice of the base year.[3] Finally, it should be noted that making cross-country comparisons is complicated by the fact that nominal exchange rates tend to fluctuate over time as well. In principle, one can correct for variation in the exchange rate, but how well this is accomplished in practice remains an open question.

[3] Some statistical agencies have introduced various 'chain-weighting' procedures to mitigate this problem.

1.5 Real GDP Across Time

Figure 1.3 plots the time path of real (i.e., corrected for inflation) per capita GDP for the United States and Canada since the first quarter of 1961.

FIGURE 1.3
Real per capita GDP
United States and Canada
1961.1 - 2003.4



The pattern of economic development for these two countries in Figure 1.3 is typical of the pattern of development observed in many industrialized countries over the last century and earlier. The most striking feature in Figure 1.3 is that real per capita income tends to grow over time. Over the last 100 years, the rate of growth in these two North American economies has averaged approximately 2% per annum.

Now, 2% per annum may not sound like a large number, but one should keep in mind that even very low rates of growth can translate into very large changes in the level of income over long periods of time. To see this, consider the 'rule of 72,' which tells us the number of years n it would take to double incomes if an economy grows at rate of g% per annum:

$$n = \frac{72}{g}.$$

Thus, an economy growing at 2% per annum would lead to a doubling of income every 36 years. In other words, we are roughly twice as rich as our predecessors who lived here in 1967; and we are four times as rich as those who lived here in 1931.

Since our current high living standards depend in large part on past growth, and since our future living standards (and those of our children) will depend on current and future growth rates, understanding the phenomenon of growth is of primary importance. The branch of macroeconomics concerned with the issue of long-run growth is called *growth theory*. A closely related branch of macroeconomics, which is concerned primarily with explaining the level and growth of incomes across countries, is called *development theory*. We will discuss theories of growth and development in the chapters ahead.

Traditionally, macroeconomics has been concerned more with the issue of 'short run' growth, or what is usually referred to as the *business cycle*. The business cycle refers to the cyclical fluctuations in GDP around its 'trend,' where trend may defined either in terms of levels or growth rates. From Figure 1.3, we see that while per capita GDP tends to rise over long periods of time, the rate of growth over short periods of time can fluctuate substantially. In fact, there appear to be (relatively brief) periods of time when the real GDP actually falls (i.e., the growth rate is negative). When the real GDP falls for two or more consecutive quarters (six months), the economy is said to be in *recession*. Figure 1.4 plots the growth rate in real per capita GDP for the United States and Canada.

FIGURE 1.4
Growth Rate in Real Per Capita GDP
(5-quarter moving average)

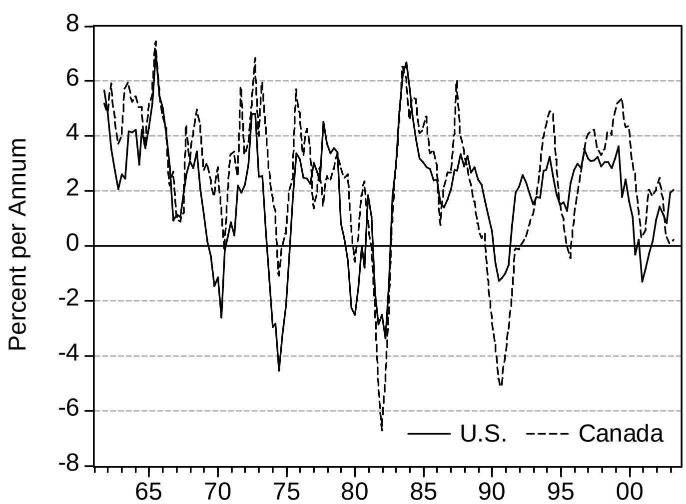


Figure 1.4 reveals that the cyclical pattern of GDP growth in the United States and Canada are similar, but not identical. In particular, note that Canada largely escaped the three significant recessions that afflicted the U.S. during the 1970s (although growth did slow down in Canada during these

episodes). Both Canada and the U.S. experienced a relatively deep recession during the early 1980s, with the recession being somewhat deeper in Canada. Both economies also experienced a recession in the early 1990s, with the recession in Canada being both deeper and more prolonged. Finally, observe that the most recent recession in the U.S. was relatively mild by historical standards. In Canada, growth fell to relatively low rates, but largely remained positive.

1.6 Schools of Thought

The reason for why aggregate economic activity fluctuates the way it does, even in relatively stable institutional environments, remains largely an unresolved puzzle. Most macroeconomists interpret the business cycle as the product of economic behavior in response to various 'shocks' that are thought to afflict economies at a high frequency. Unfortunately, these high-frequency shocks (if they exist at all) are difficult to measure directly, leading to much debate concerning the ultimate source of the business cycle. Furthermore, economists are often divided on how they view an economy reacting to any given shock. Thus there are many different schools of thought that are distinguished by which shocks they choose to emphasize and/or in their explanation of how the economy reacts to any given shock.

There are two broad strands of thinking (each with its own many variations) in terms of understanding the business cycle. The first strand, which I will refer to as the 'conventional wisdom,' owes its intellectual debt primarily to the work of John Maynard Keynes (1936), whose views on the business cycle were likely shaped to a large extent by the experience of the Great Depression. Almost every principles course in macroeconomics is taught according to the Keynesian perspective, which goes a long way to explaining why this view has become the conventional one among market analysts, central bankers, and policymakers in general. In academic circles, certain elements of this view live on in the work of the *New-Keynesian* school of thought.

According to conventional wisdom, the long-run trend rate of growth in real per capita GDP is considered to be more or less constant (e.g., 2% per annum). The level of GDP consistent with this trend is sometimes referred to as *potential GDP*. Fluctuations in GDP around trend are thought to be induced by various 'aggregate demand' shocks, for example, an unexplained sudden surge in desired spending on the part of the consumer, the business sector, the government sector, or the foreign sector. At a deeper level, the private sector spending shocks are commonly thought to be the result of 'irrational' swings in consumer and business sector 'sentiment,' or what Keynes (1936) referred to as 'animal spirits.' Any given shock influences market outcomes (like the GDP and employment) adversely, primarily because private markets are viewed as being subject to a host of various 'imperfections.' These imperfections are usually thought to take the form of 'sticky' prices and/or credit market imperfections arising from imperfect information in financial markets.

The conventional view asserts that in an ideal world, the economy's actual GDP would always be equal to its potential GDP. In other words, the business cycle is largely viewed as evidence of a malfunctioning market system. Given this interpretation of the cycle, it should come as no surprise that this view also advocates the use of various government stabilization policies (active monetary and fiscal policy) to mitigate the adverse consequences of the cycle.

The second strand of thinking is the *neoclassical perspective*. This school of thought is closer in spirit to the views of Joseph Schumpeter (1939). In academic circles, the neoclassical perspective is embodied in the work of a school of thought called *Real Business Cycle Theory*.

According to the neoclassical view, the distinction between 'growth' and 'cycles' is largely an artificial one. Almost everyone agrees that long-run growth is the product of technological advancement. But unlike the New-Keynesian school, which views trend growth as being relatively stable, the neoclassical view is that there is no God-given reason to believe that the process of technological advancement proceeds in such a 'smooth' manner. Indeed, it seems more reasonable to suppose that new technologies appear in 'clusters' over time. These 'technology shocks' may cause fluctuations in the trend rate of growth through what Schumpeter called a process of 'creative destruction.' That is, technological advancements that ultimately lead to higher productivity may, in the short run, induce cyclical adjustments as the economy 'restructures' (i.e., as resources flow from declining sectors to expanding sectors).

As with the conventional wisdom, the neoclassical view admits that sudden changes in private sector expectations may lead to sudden changes in desired consumer and business sector spending. But unlike the conventional wisdom, these changes are interpreted as reflecting the 'rational' behavior of private sector decision-makers in *response* to perceived real changes in underlying economic fundamentals (i.e., technology shocks). In other words, changes in market sentiment are the *result* and not the *cause* of the business cycle.

According to the neoclassical view, the business cycle is an unfortunate but largely unavoidable product of the process of economic development. Given this interpretation of the cycle, it should come as no surprise to learn that the policy implication here is that government attempts to stabilize the cycle are likely to do more harm than good.

In the chapters to come, we will explore the relative merits and shortcomings of each of these perspectives. For the student, I recommend keeping an open mind with respect to each perspective, since there is likely an element of truth in each of these diverse views.

1.7 Problems

1. While Americans constitute a relatively small fraction of the world's population (less than 5%), they spend approximately 20% of the world's income. This fact is sometimes used as evidence of American 'greed.' Provide a different interpretation of this fact based on your knowledge of the relationship between aggregate expenditure and output.

2. One often reads how much of the world's population subsists on 'less than $1 a day.' Explain why such reports grossly underestimate the level of per capita income actually produced in the world's underdeveloped economies (see Appendix 1.A).

3. The residential capital stock produces a flow of output in the form of 'shelter services.' When a landlord rents out his or her residence, the value of this output is measured by the rental income generated by the asset. However, many residences are 'owner-occupied' so that no direct measure of rental income is available (for example, if you own your own home, you do not typically write out a rental cheque to yourself each month). Nevertheless, statisticians attempt to impute a number to the value of output generated by owner-occupied housing. Explain how this might be done.

4. Your father keeps a vegetable garden in the backyard, the output of which is consumed by household members. Should the value of this output be counted toward an economy's GDP? If so, how might a statistician measure the value of such output from a practical standpoint? Explain any similarity and differences between this example and the previous example of owner-occupied housing.

5. Explain why 'overpaid' government employees will lead to an overstatement of GDP, whereas 'overpaid' private sector employees will not (see Appendix 1.A).

6. Consider an economy that produces bananas and bread. You have the following measurements:

	Bananas	Bread
Quantity (2003)	100	50
Price (2003)	$0.50	$1.00
Quantity (2004)	110	60
Price (2004)	$0.50	$1.50

   (a) Compute the GDP in each year using current prices. Compute the growth rate in nominal GDP.

   (b) Compute the real GDP in each year first using 2003 as the base year and then using 2004 as the base year. How does the rate of growth in real GDP depend on which base year is used? Explain. (Hint: note that the price of bread relative to bananas has increased).

7. Consider two economies A and B that each have a real per capita GDP equal to \$1,000 in the year 1900. Suppose that economy A grows at 2% per annum, while economy B grows at 1.5% per annum. The difference in growth rates does not seem very large, but compute the GDP in these two economies for the year 2000. In percentage terms, how much higher is the real GDP in economy A compared to economy B?

1.8 References

1. Keynes, John M. (1936). *The General Theory of Employment*, Interest and Money, MacMillan, Cambridge University Press.

2. Schumpeter, Joseph A. (1939). *Business Cycles: A Theoretical, Historical and Statistical Analysis of the Capitalist Process*, New York, McGraw-Hill.

1.A Measured GDP: Some Caveats

Have you ever wondered how GDP figures are calculated in practice? Probably not...but let me tell you anyway.

In principle, the GDP is calculated either by summing up incomes or expenditures over some specified period of time (e.g., month, quarter, year). This sounds simple in principle–but how is it done in practice? Does the government or some other organization keep a running tab on everybody's income and expenditures at all times? No, this is not how it works.

In practice, data is collected from a variety of sources using a variety of methods. For example, government statistical agencies may have access to payroll data or personal and business income tax forms and use this information to construct an estimate of income. As well, these same statistical agencies may perform regular surveys of randomly selected establishments (within a variety of sectors) to gather sales data from which to form an estimate of expenditure. Thus, it is important to keep in mind that available GDP numbers are just *estimates* (which are often revised over time), whose quality depends in part on the resources that are spent in collecting information, the information that is available, and on the methodology used in constructing estimates. It is not clear that any of these factors remain constant over time or are similar across countries.

Another thing to keep in mind is that measuring the value of income and expenditure in the way just described largely limits us to measurements of income and expenditure (and hence, the production of output) that occur in 'formal' markets. For present purposes, one can think of a formal market as a place where: [1] output exchanges for money at an observable price; and [2] the value of what is exchanged is observable by some third party (i.e., the government or some statistical agency).

It is likely that most of the marketable output produced in the so-called developed world is exchanged in formal markets, thanks largely to an extensive market system and revenue-hungry governments eager to tax everything they can (in the process keeping records of the value of what is taxed). In many economies, however, a significant fraction of marketable output is likely exchanged in markets with observable money prices, but where the value of what is exchanged is 'hidden' (and hence, not measurable). The reason for hiding the value of such exchanges is often motivated by a desire to evade taxes or because the exchange itself is legally prohibited (e.g., think of the marijuana industry in British Columbia or the cocaine industry in Colombia). The output produced in the so-called 'underground economy' should, in principle, be counted as part of an economy's GDP, but is not owing to the obvious problem of getting individuals to reveal their underground activities.

A large fraction of the marketable output produced in the so-called underdeveloped world is likely exchanged in informal markets. The reason for this is twofold. First, one thing that characterizes underdeveloped economies is lack of formal market structures, particularly in rural areas (where output is largely consumed by the household, or perhaps bartered in small and informal transactions). The lack of local market prices thus requires statistical agencies to impute market value, which are calculations that require resources that may not be available. Second, many rural regions in the underdeveloped world operate largely in the absence of any (federal) government intervention (taxation). Since tax information constitutes an important source of data for estimating income, the lack of any federal-level record-keeping system outside of urban areas may severely hinder the collection of any relevant data. These measurement problems should especially be kept in mind when attempting to make cross-country comparisons of GDP (especially across developed and less-developed economies).

Finally, a word needs to be said about an asymmetry that exists in the measurement of private versus public sector activity. Consider the following example. The private sector consists of a household sector and a business sector. The household sector generates labor income and the business sector generates capital income. Members of the household sector are employed in either the business sector or the government sector. Let W_P denote wage income from those employed in the business sector and W_G denote wage income from those employed in the government sector. The household sector also pays taxes T, so that total after-tax household earnings are given by:

$$W_P + W_G - T.$$

The business sector produces and sells output (at market prices), which generates revenue Y_P. Capital income is therefore given by:

$$Y_P - W_P.$$

Adding up the two equations above yields us the after-tax income of the private sector:

$$Y_P + W_G - T.$$

The government sector produces output Y_G. Measuring the value of Y_G, however, is problematic since most of what a government produces is not sold at market prices (it is largely given away). Nevertheless, one could impute a value for Y_G. Having done so, government sector income is given by:

$$Y_G + T - W_G.$$

If we add up private-sector and government-sector income (the previous two equations), we arrive at a figure for total income; i.e., $Y = Y_P + Y_G$.

The key question here is how to impute a value for Y_G? The way this is done in practice is to assume:

$$Y_G = W_G.$$

In other words, it is assumed that the value of what is produced by government employees is equal to what the government pays for their services.

To see the potential problem of this method of imputation, consider an extreme case in which government employees produce nothing of value so that in truth $Y_G = 0$ and $Y = Y_P$. In this case, the government wage bill W_G in fact represents a *transfer* of income from one part of the population to another; i.e., it does not represent the (market) value of newly produced goods and services. In this case, imputing the value $Y_G = W_G$ overstates the value of government-produced output and will therefore overstate an economy's GDP.

Chapter 2

Basic Neoclassical Theory

2.1 Introduction

In this chapter, we develop a simple theory (based on the neoclassical perspective) that is designed to explain the determination of output and employment (hours worked). The object is to construct a *model economy*, populated by individuals that make certain types of decisions to achieve some specified goal. The decisions that people make are subject to a number of constraints so that inevitably, achieving any given goal involves a number of trade-offs. In many respects, the theory developed here is too simple and suffers from a number of shortcomings. Nevertheless, it will be useful to study the model, since it serves as a good starting point and can be extended in a number of dimensions as the need arises.

For the time being, we will focus on the output of consumer goods and services (hence, ignoring the production of new capital goods or investment); i.e., so that $I \equiv 0$. For simplicity, we will focus on an economy in which labor is the only factor of production (Appendix 2.A extends the model to allow for the existence of a productive capital stock). For the moment, we will also abstract from the government sector, so that $G \equiv 0$. Finally, we consider the case of a closed economy (no international trade in goods or financial assets), so that $NX \equiv 0$. From our knowledge of the income and expenditure identities, it follows that in this simple world, $C \equiv Y \equiv L$. In other words, all output is in the form of consumer goods purchased by the private sector and all (claims to) output are paid out to labor.

A basic outline of the neoclassical model is as follows. First, it is assumed that individuals in the economy have preferences defined over consumer goods and services so that there is a demand for consumption. Second, individuals also have preferences defined over a number of nonmarket goods and services, that are produced in the home sector (e.g., leisure). Third, individuals are endowed with a fixed amount of time that they can allocate either to the labor market or the home sector. Time spent in the labor market is useful for the purpose of earning wage income, which can be spent on consumption. On the other hand, time spent in the labor market necessarily means that less time can be spent in other valued activities (e.g., home production or leisure). Hence individuals face a trade-off: more hours spent working imply a higher material living standard, but less in the way of home production (which is not counted as GDP). A key variable that in part determines the relative returns to these two activities is the real wage rate (the purchasing power of a unit of labor).

The production of consumer goods and services is organized by firms in the business sector. These firms have access to a production technology that transforms labor services into final output. Firms are interested in maximizing the return to their operations (profit). Firms also face a trade-off: Hiring more labor allows them to produce more output, but increases their costs (the wage bill). The key variables that determine the demand for labor are: (a) the productivity of labor; and (b) the real wage rate (labor cost).

The real wage is determined by the interaction of individuals in the household sector and firms in the business sector. In a competitive economy, the real wage will be determined by (among other things) the productivity of labor. The productivity of labor is determined largely by the existing structure of technology. Hence, fluctuations in productivity (brought about by technology shocks) may induce fluctuations in the supply and demand for labor, leading to a business cycle.

2.2 The Basic Model

The so-called basic model developed here contains two simplifying assumptions. First, the model is 'static' in nature. The word 'static' should not be taken to mean that the model is free of any concept of time. What it means is that the decisions that are modeled here have no intertemporal dimension. In particular, choices that concern decisions over how much to save or invest are abstracted from. This abstraction is made primarily for simplicity and pedagogy; in later chapters, the model will be extended to 'dynamic' settings. The restriction to static decision-making allows us, for the time-being, to focus on *intratemporal* decisions (such as the division of time across competing uses). As such, one can interpret the economy as a sequence $t = 1, 2, 3, ..., \infty$ of static outcomes.

The second abstraction involves the assumption of 'representative agencies.' Literally, what this means is that all households, firms and governments are assumed to be identical. This assumption captures the idea that individual agencies share many key characteristics (e.g., the assumption that more is preferred to less) and it is these key characteristics that we choose to emphasize. Again, this assumption is made partly for pedagogical reasons and partly because the issues that concern us here are unlikely to depend critically on the fact that individuals and firms obviously differ along many dimensions. We are not, for example, currently interested in the issue of income distribution. It should be kept in mind, however, that the neoclassical model can be (and has been) extended to accommodate heterogeneous decision-makers.

2.2.1 The Household Sector

Imagine an economy with (identical) households that each contain a large number (technically, a continuum) of individuals. The welfare of each household is assumed to depend on two things: (1) a basket of consumer goods and services (consumption); and (2) a basket of home-produced goods and services (leisure). Let c denote consumption and let l denote leisure. Note that the value of home-produced output (leisure) is not counted as a part of the GDP.

How do households value different combinations of consumption and leisure? We assume that households are able to rank different combinations of (c, l) according to a utility function $u(c, l)$. The utility function is just a mathematical way of representing household preferences. For example, consider two 'allocations' (c_A, l_A) and (c_B, l_B). If $u(c_A, l_A) > u(c_B, l_B)$, then the household prefers allocation A to allocation B; and vice-versa if $u(c_A, l_A) < u(c_B, l_B)$. If $u(c_A, l_A) = u(c_B, l_B)$, then the household is indifferent between the two allocations. We will assume that it is the goal of each household to act in a way that allows them to achieve the highest possible utility. In other words, households are assumed to do the best they can according to their preferences (this is sometimes referred to as maximizing behavior).

It makes sense to suppose that households generally prefer more of c and ℓ to less, so that $u(c, l)$ is *increasing* in both c and l. It might also make sense to suppose that the function $u(c, l)$ displays *diminishing marginal utility* in both c and l. In other words, one extra unit of either c or l means a lot less to me if I am currently enjoying high levels of c and l. Conversely, one extra unit of either c or l would mean a lot more to me if I am currently enjoying low levels of c and l.

Now, let us fix a utility number at some arbitrary value; i.e., u_0. Then, consider the expression:

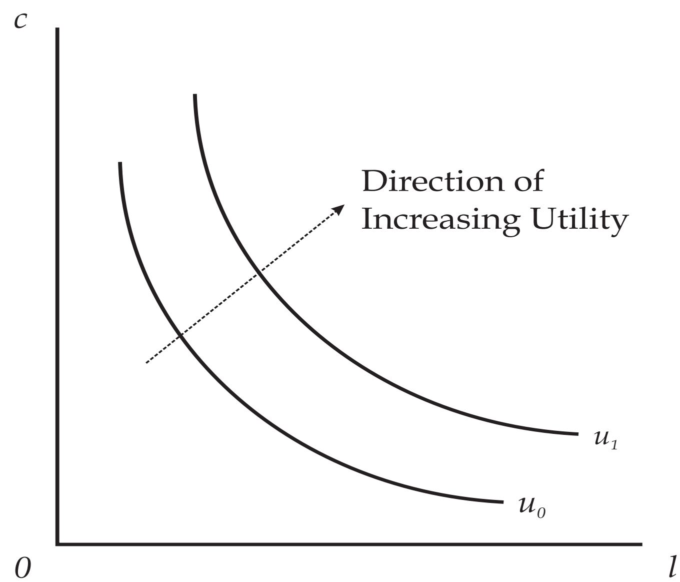
$$u_0 = u(c, l). \tag{2.1}$$

This expression tells us all the different combinations of c and l that generate the utility rank u_0. In other words, the household is by definition *indifferent* between all the combinations of c and l that satisfy equation (2.1). Not surprisingly, economists call such combinations an *indifference curve.*

Definition: An indifference curve plots all the set of allocations that yield the same utility rank.

If the utility function is increasing in both c and l, and if preferences are such that there is diminishing marginal utility in both c and ℓ, then indifference curves have the properties that are displayed in Figure 2.1, where two indifference curves are displayed with $u_1 > u_0$.

FIGURE 2.1
Indifference Curves



Households are assumed to have *transitive* preferences. That is, if a household prefers (c_1, l_1) to (c_2, l_2) and also prefers (c_2, l_2) to (c_3, l_3), then it is also true that the household prefers (c_1, l_1) to (c_3, l_3). The transitivity of preferences implies the following important fact:

Fact: *If preferences are transitive, then indifference curves can never cross.*

Keep in mind that this fact applies to a *given* utility function. If preferences were to change, then the indifference curves associated with the original preferences may cross those indifference curves associated with the new preferences. Likewise, the indifference curves associated with two different households may also cross, without violating the assumption of transitivity. Ask your instructor to elaborate on this point if you are confused.

An important concept associated with preferences is the *marginal rate of substitution*, or MRS for short. The definition is as follows:

Definition: The *marginal rate of substitution* (MRS) between any two goods is defined as the (absolute value of the) slope of an indifference curve at any allocation.

The MRS has an important economic interpretation. In particular, it measures the household's *relative valuation* of any two goods in question (in this case, consumption and leisure). For example, consider some allocation (c_0, l_0), which is given a utility rank $u_0 = u(c_0, l_0)$. How can we use this information to measure a household's relative valuation of consumption and leisure? Imagine taking away a small bit Δ_l of leisure from this household. Then clearly, $u(c_0, l_0 - \Delta_l) < u_0$. Now, we can ask the question: How much extra consumption Δ_c would we have to compensate this household such that they are not made any worse off? The answer to this question is given by the Δ_c that satisfies the following condition:

$$u_0 = u(c_0 + \Delta_c, l_0 - \Delta_l).$$

For a very small Δ_l, the number Δ_c/Δ_l gives us the slope of the indifference curve in the neighborhood of the allocation (c_0, l_0). It also tells us how much this household values consumption relative to leisure; i.e., if Δ_c/Δ_l is large, then leisure is valued highly (one would have to give a lot of extra consumption to compensate for a small drop in leisure). The converse holds true if Δ_c/Δ_l is a small number.

Before proceeding, it may be useful to ask why we (as theorists) should be interested in modeling household preferences in the first place. There are at least two important reasons for doing so. First, one of our goals is to try to predict household behavior. In order to predict how households might react to any given change in the economic environment, one presumably needs to have some idea as to what is motivating their behavior in the first place. By specifying the objective (i.e., the utility function) of the household explicitly, we can use this information to help us predict household behavior. Note that this remains true even if we do not know the exact form of the utility function $u(c, l)$. All we really need to know (at least, for making qualitative predictions) are the general properties of the utility function (e.g., more is preferred to less, etc.). Second, to the extent that policymakers are concerned with implementing policies that improve the welfare of individuals, understanding how different policies affect household utility (a natural measure of economic welfare) is presumably important.

Now that we have modeled the household objective, $u(c, l)$, we must now turn to the question of what constrains household decision-making. Households are endowed with a fixed amount of time, which we can measure in units of either hours or individuals (assuming that each individual has one unit of time). Since the total amount of available time is fixed, we are free to normalize this number to unity. Likewise, since the size of the household is also fixed, let us normalize this number to unity as well.

Households have two competing uses for their time: work (n) and leisure (l), so that:

$$n + l = 1. \tag{2.2}$$

Since the total amount of time and household size have been normalized to unity, we can interpret n as either the fraction of time that the household devotes to work or the fraction of household members that are sent to work at any given date.

Now, let w denote the consumer goods that can be purchased with one unit of labor. The variable w is referred to as the *real wage*. For now, let us simply assume that $w > 0$ is some arbitrary number beyond the control of any individual household (i.e., the household views the market wage as exogenous). Then consumer spending for an individual household is restricted by the following equation:[1]

$$c = wn.$$

By combining the time constraint (2.2) with the budget constraint above, we see that household choices of (c, l) are in fact constrained by the equation:

$$c = w - wl. \tag{2.3}$$

This constraint makes it clear that an increase in l necessarily entails a reduction in material living standards c.

Before proceeding, a remark is in order. Note that the 'money' that workers get paid is in the form of a privately-issued claim against the output to be produced in the business sector. As such, this 'money' resembles a coupon issued by the firm that is redeemable for merchandise produced by the firm.[2]

We are now ready to state the choice problem facing the household. The household desires an allocation (c, l) that maximizes utility $u(c, l)$. However, the choice of this allocation (c, l) must respect the budget constraint (2.3). In mathematical terms, the choice problem can be stated as:

Choose (c, l) to maximize $u(c, l)$ subject to: $c = w - wl$.

Let us denote the optimal choice (i.e., the solution to the choice problem above) as (c^D, l^D), where $c^D(w)$ can be thought of as consumer demand and $l^D(w)$ can be thought of as the demand for leisure (home production). In terms of a diagram, the optimal choice is displayed in Figure 2.2 as allocation A.

[1] This equation anticipates that, in equilibrium, non-labor income will be equal to zero. This result follows from the fact that we have assumed competitive firms operating a technology that utilizes a single input (labor). When there is more than one factor of production, the budget constraint must be modified accordingly; i.e., see Appendix 2.A.

[2] For example, in Canada, the firm *Canadian Tire* issues its own money redeemable in merchandise. Likewise, many other firms issue coupons (e.g., gas coupons) redeemable in output. The basic neoclassical model assumes that all money takes this form; i.e., there is no role for a government-issued payment instrument. The subject of money is taken up in later chapters.

FIGURE 2.2
Household Choice

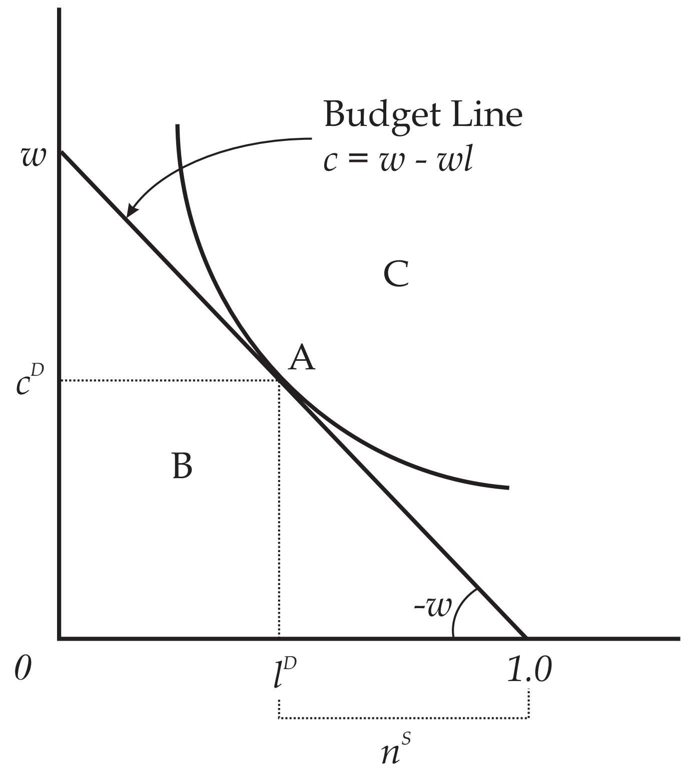


Figure 2.2 contains several pieces of information. First note that the budget line (the combinations of c and l that exhaust the available budget) is linear, with a slope equal to $-w$ and a y-intercept equal to w. The y-intercept indicates the maximum amount of consumption that is budget feasible, given the prevailing real wage w. In principle, allocations such as point B are also budget feasible, but they are not optimal. That is, allocation A is preferred to B and is affordable. An allocation like C is preferred to A, but note that allocation C is not affordable. The best that the household can do, given the prevailing wage w, is to choose an allocation like A.

As it turns out, we can describe the optimal allocation mathematically. In particular, one can prove that only allocation A satisfies the following two conditions at the same time:

$$\begin{aligned} MRS(c^D, l^D) &= w; \\ c^D &= w - wl^D. \end{aligned} \tag{2.4}$$

The first condition states that, at the optimal allocation, the slope of the indifference curve must equal the slope of the budget line. The second condition states that the optimal allocation must lie on the budget line. Only the allocation at point A satisfies these two conditions simultaneously.

- **Exercise 2.1.** Using a diagram similar to Figure 2.2, identify an allocation that satisfies $MRS = w$, but is not on the budget line. Can such an allocation be optimal? Now identify an allocation that is on the budget line, but where $MRS \neq w$. Can such an allocation be optimal? Explain.

Finally, observe that this theory of household choice implies a theory of labor supply. In particular, once we know l^D, then we can use the time constraint to infer that the desired household labor supply is given by $n^S = 1 - l^D$. Thus, the solution to the household's choice problem consists of a set of functions: $c^D(w)$, $l^D(w)$, and $n^S(w)$.

Substitution and Wealth Effects Following a Wage Change

Figure 2.3 depicts how a household's desired behavior may change with an increase in the return to labor. Let allocation A in Figure 2.3 depict desired behavior for a low real wage, w_L. Now, imagine that the real wage rises to $w_H > w_L$. Figure 2.3 shows that the household may respond in three general ways, which are represented by the allocations B,C, and D.

An increase in the real wage has two effects on the household budget. First, the price of leisure (relative to consumption) increases. Second, household wealth (measured in units of output) increases. These two effects can be seen in the budget constraint (2.3), which one can rewrite as:

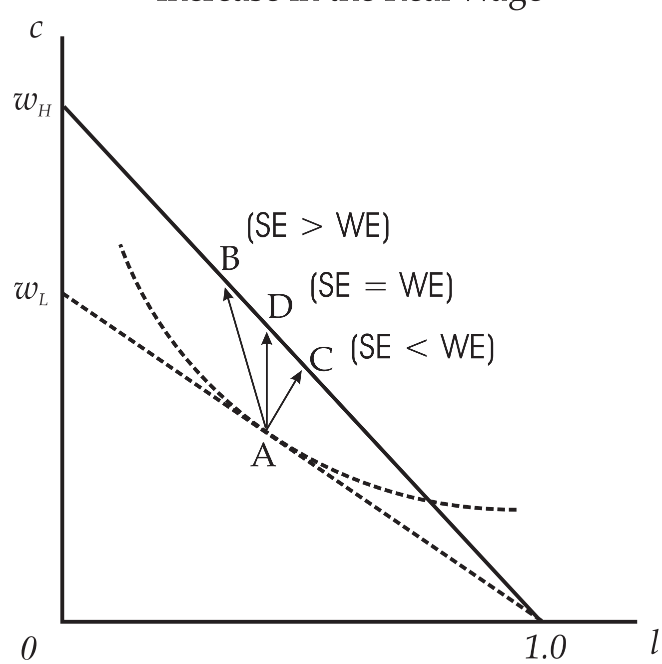
$$c + wl = w.$$

The right hand side of this equation represents household wealth, measured in units of consumption (i.e., the y-intercept in Figure 2.3). Thus a change in w_t induces what is called a *wealth effect* (WE). The left hand side represents the household's expenditure on consumption and leisure. The price of leisure (measured in units of foregone consumption) is w; i.e., this is the slope of the budget line. Since a change in w also changes the relative price of consumption and leisure, it will induce what is called a *substitution effect* (SE).

From Figure 2.3, we see that an increase in the real wage is predicted to increase consumer demand. This happens because: (1) the household is wealthier (and so can afford more consumption); and (2) the price of consumption falls (relative to leisure), inducing the household to substitute away from leisure and into consumption. Thus, both wealth and substitution effects work to increase consumer demand.

Figure 2.3 also suggests that the demand for leisure (the supply of labor) may *either* increase or decrease following an increase in the real wage. That is, since wealth has increased, the household can now afford to purchase more leisure (so that labor supply falls). On the other hand, since leisure is more expensive (the return to work is higher), the household may wish to purchase less leisure (so that labor supply rises). If this substitution effect dominates the wealth effect, then the household will choose an allocation like B in Figure 2.3. If the wealth effect dominates the substitution effect, then the household will choose an allocation like C. If these two effects exactly cancel, then the household will choose an allocation like D (i.e., the supply of labor does not change in response to an increase in the real wage). But which ever case occurs, we can conclude that the household is made better off (i.e., they will achieve a higher indifference curve).

FIGURE 2.3
Household Response to an
Increase in the Real Wage



- **Exercise 2.2.** Consider the utility function $u(c, l) = \ln c + \lambda \ln l$, where $\lambda \geq 0$ is a preference parameter that measures how strongly a household feels about consuming home production (leisure). For these preferences, we have $MRS = \lambda c/l$. Using the conditions in (2.4), solve for the household's labor supply function. How does labor supply depend on the real wage here? Explain. Suppose now that preferences are such that the MRS is given by $MRS = (c/l)^{1/2}$. How does labor supply depend on the real wage? Explain.

- **Exercise 2.3.** Consider two household's that have preferences as in the exercise above, but where preferences are distinguished by different values for λ; i.e., $\lambda_H > \lambda_L$. Using a diagram similar to Figure 2.2, depict the different choices made by each household. Explain. (Hint: the indifference curves will cross).

2.2.2 The Business Sector

Our model economy is populated by a number of (identical) business agencies that operate a production technology that transforms labor services (n) into output (in the form of consumer goods and services) (y). We assume that this production technology takes a very simple form:

$$y = zn;$$

where $z > 0$ is a parameter that indexes the efficiency of the production process. We assume that z is determined by forces that are beyond the control of any individual or firm (i.e., z is exogenous to the model).

In order to hire workers, each firm must pay its workers the market wage w. Again, remember that the assumption here is that firms can create the 'money' they need by issuing coupons redeemable in output. Let d denote the profit (measured in units of output) generated by a firm; i.e.,

$$d = (z - w)n. \tag{2.5}$$

Thus, the choice problem for a firm boils down to choosing an appropriate labor force n; i.e.,

> Choose (n) to maximize $(z - w)n$ subject to $0 \leq n \leq 1$.

The solution to this choice problem, denoted n^D (the labor demand function), is very simple and depends only on (z, w). In particular, we have:

$$n^D = \begin{cases} 0 & \text{if} \quad z < w; \\ n & \text{if} \quad z = w; \\ 1 & \text{if} \quad z > w; \end{cases}$$

where n in the expression above is any number in between 0 and 1. In words, if the return to labor (z) is less than the cost of labor (w), then the firm will demand no workers. On the other hand, if the return to labor exceeds the cost of labor, then the firm will want to hire all the labor it can. If the return to labor equals the cost of labor, then the firm is indifferent with respect to its choice of employment (the demand for labor is said to be *indeterminate* in this case). With the demand for labor determined in this way, the supply of output is simply given by $y^S = zn^D$.

Notice that the demand for labor depends on both w and z, so that we can write $n^D(w, z)$. Labor demand is (weakly) decreasing in w. That is, suppose that $z > w$ so that labor demand is very high. Now imagine increasing w higher and higher. Eventually, labor demand will fall to zero. The demand for labor is also (weakly) increasing in z. To see this, suppose that initially $z < w$. Now imagine increasing z higher and higher. Eventually, labor demand will be equal 1. Thus, anything that leads to an improvement in the efficiency of the production technology will generally serve to increase the demand for labor.

2.2.3 General Equilibrium

So far, we have said nothing about how the real wage (the relative price of output and leisure) is determined. In describing the choice problem of households and firms, we assumed that the real wage was beyond the control of any individual household or firm. This assumption can be justified by the fact that, in a competitive economy, individuals are small relative to the entire economy, so that individual decisions are unlikely to influence market prices. But market prices do not generally fall out of thin air. It makes sense to think of prices as ultimately being determined by aggregate supply and aggregate demand conditions in the market place. So, it is now time to bring households and firms together and describe how they interact in the market place. The outcome of this interaction is called a *general equilibrium*.

The economy's general equilibrium is defined as an allocation (c^*, y^*, n^*, l^*) and a price system (w^*) such that the following is true:

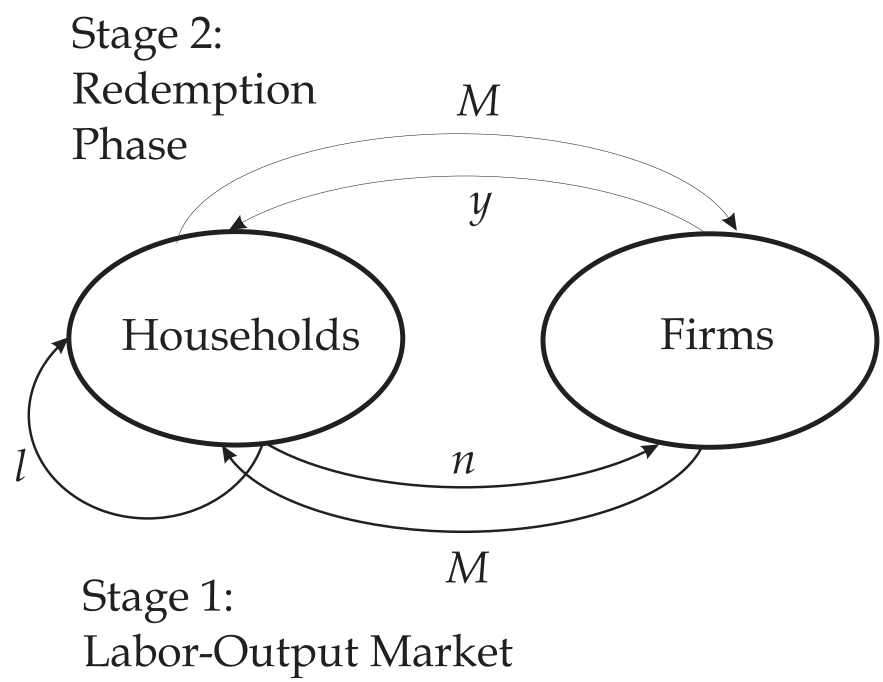
1. Given w^*, the allocation (c^*, n^*, l^*) maximizes utility subject to the budget constraint [households are doing the best they can];
2. Given (w^*, z), the allocation (y^*, n^*) maximizes profit [firms are doing the best they can];
3. The price system (w^*) clears the market [$n^S(w^*) = n^D(w^*, z)$ or $c^D(w^*) = y^S(w^*, z)$].

In words, the general equilibrium concept is asking us to interpret the world as a situation in which all of its actors are trying to do the best they can (subject to their constraints) in competitive markets. Observed prices (equilibrium prices) are likewise interpreted to be those prices that are consistent with the optimizing actions of all individuals taken together.

Before we examine the characteristics of the general equilibrium, it is useful to summarize the pattern of exchanges that are imagined to occur in each period; i.e., see Figure 2.4. One can imagine that each period is divided into two stages.

In the first stage, workers supply their labor (n) to firms in exchange for coupons (M) redeemable for y units of output. The real GDI at this stage is given by y. In the second stage (after production has occurred), households take their coupons (M) and redeem them for output (y). Since M represents a claim against y, the real GDE at this stage is given by y. And since firms actually produce y, the real GDP is given by y as well.

FIGURE 2.4
Pattern of Exchange

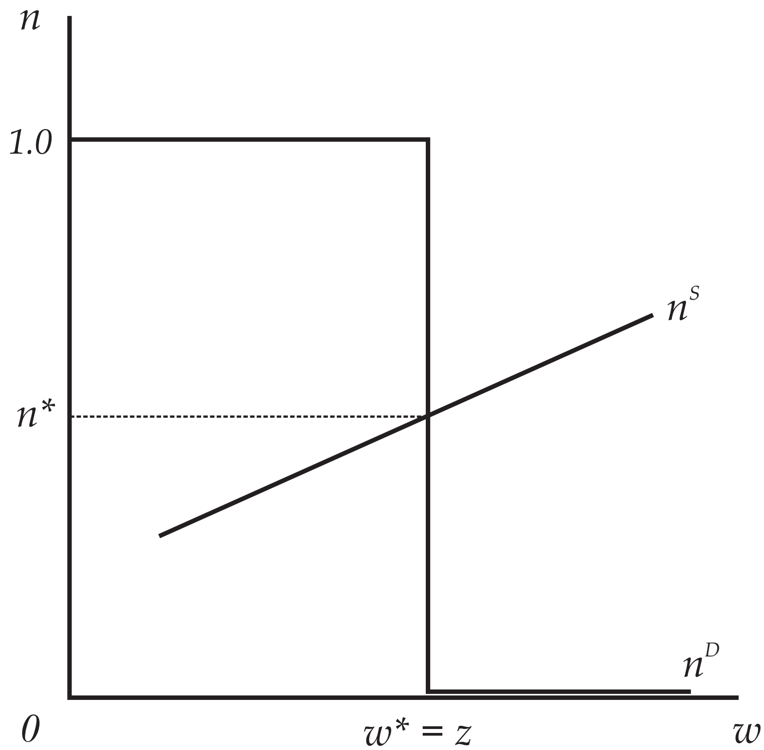


Now let us proceed to describe the general equilibrium in more detail. From the definition of equilibrium, real wage will satisfy the labor market clearing condition:

$$n^S(w^*) = n^D(w^*, z).$$

If household preferences are such that zero levels of consumption or leisure are extremely undesirable, then the equilibrium will look something like that depicted in Figure 2.5.

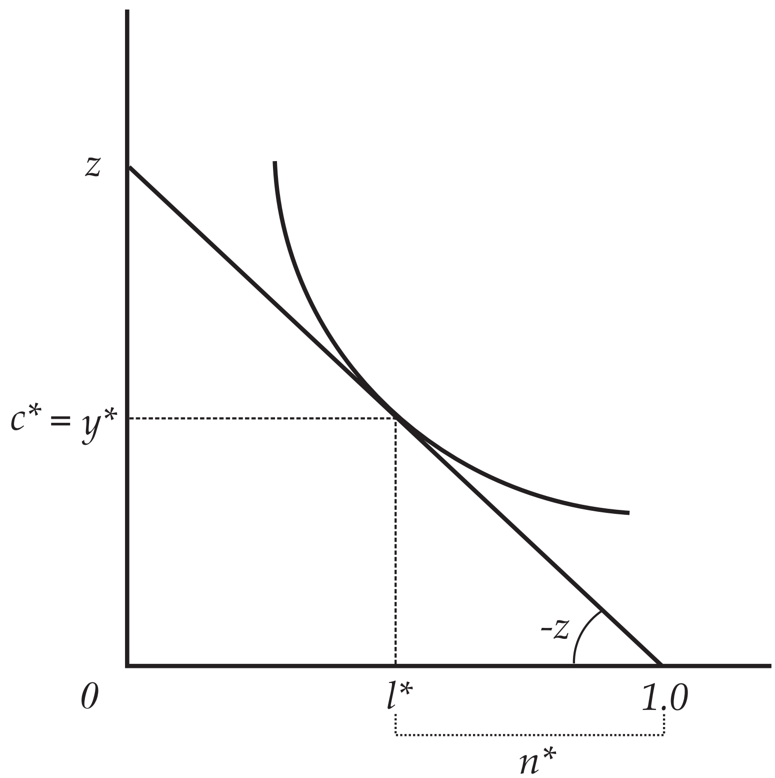
FIGURE 2.5
Equilibrium in the Labor Market



Notice that the upward sloping labor supply function reflects an underlying assumption about the relative strength of the substitution and wealth effects associated with changes in the real wage; i.e., here we are assuming that SE > WE. The peculiar shape of the labor demand function follows from our assumption that the production function is linear in employment. If instead we assumed a diminishing marginal product of labor (the standard assumption), the labor demand function would be decreasing smoothly in the real wage. Such an extension is considered in Appendix 2.A.

In general, the equilibrium real wage is determined by both labor supply and demand (as in Appendix 2.A). However, in our simplified model (featuring a linear production function), we can deduce the equilibrium real wage solely from labor demand. In particular, recall that the firm's profit function is given by $d = (z - w)n$. For n^* to be strictly between 0 and 1, it must be the case that $w^* = z$ (so that $d^* = 0$). That is, the real wage must adjust to drive profits to zero so that the demand for labor is indeterminate. With w^* determined in this way, the equilibrium level of employment is then determined entirely by the labor supply function; i.e., $n^* = n^S(w^*)$. The general equilibrium allocation and equilibrium real wage is depicted in Figure 2.6.

FIGURE 2.6
General Equilibrium



Observe that in this simple model, the *equilibrium* budget constraint just happens to lie on top of the PPF. For this reason, it may be easy to confuse the two, but you should always remember that the budget constraint and the PPF are not equivalent concepts.[3]

Thus, our theory makes a prediction over various real quantities (c^*, y^*, n^*, l^*) and the real wage w^*. These 'starred' variables are referred to as the model's *endogenous variables*; i.e., these are the objects that the theory is designed to explain. Notice that the theory here (as with any theory) contains variables that have no explanation (as far as the theory is concerned). These variables are treated as 'God-given' and are labelled *exogenous variables*. In the theory developed above, the key exogenous variables are z (technology) and u (preferences). Thus, the theory is designed to explain how its endogenous variables

[3] For example, see the more general model developed in Appendix 2.1.

react to any particular (unexplained) change in the set of exogenous variables.

- **Exercise 2.4.** Consider the general equilibrium allocation depicted in Figure 2.6. Is the real GDP maximized at this allocation? If not, which allocation does maximize GDP? Would such an allocation also maximize economic welfare? If not, which allocation does maximize economic welfare? What are the policy implications of what you have learned here?

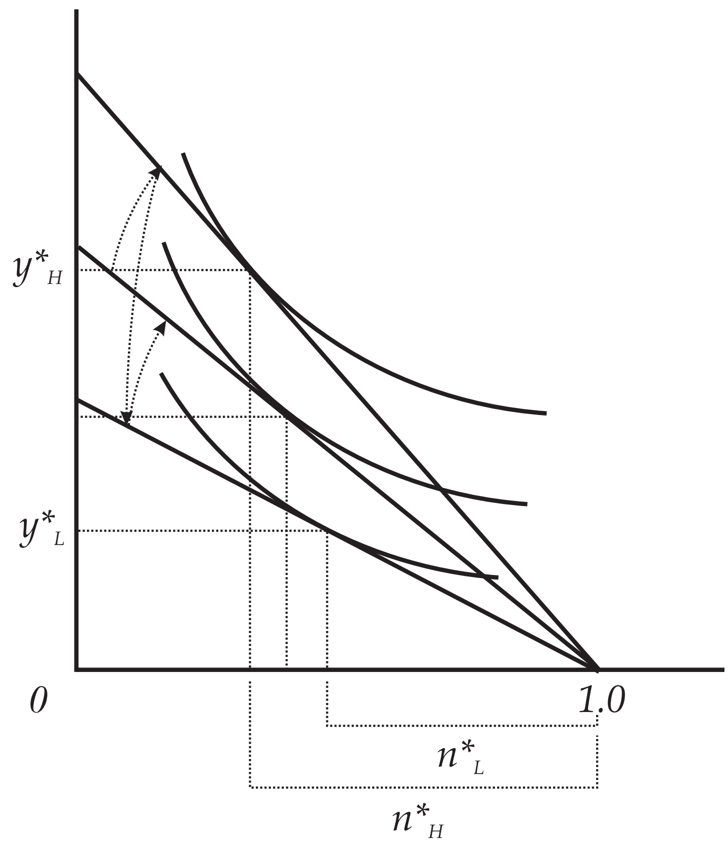
2.3 Real Business Cycles

Many economists in the past have pointed out that the process of technological development itself may be largely responsible for the business cycle.[4] The basic idea is that we have no reason to believe *a priori* that productivity should grow in perfectly smooth manner. Perhaps productivity growth has a relatively smooth trend, but there are likely to be transitory fluctuations in productivity around trend as the economy grows.

To capture this idea in our model, let us suppose that the productivity parameter z fluctuates around some given 'trend' level. Figure 2.7 depicts how the equilibrium allocation will fluctuate (assuming that SE > WE on labor supply) across three different productivity levels: $z_H > z_M > z_L$.

[4] Classic examples include: Schumpeter (1942), Kydland and Prescott (1982), and Long and Plosser (1983).

FIGURE 2.7
Business Cycles: Productivity Shocks



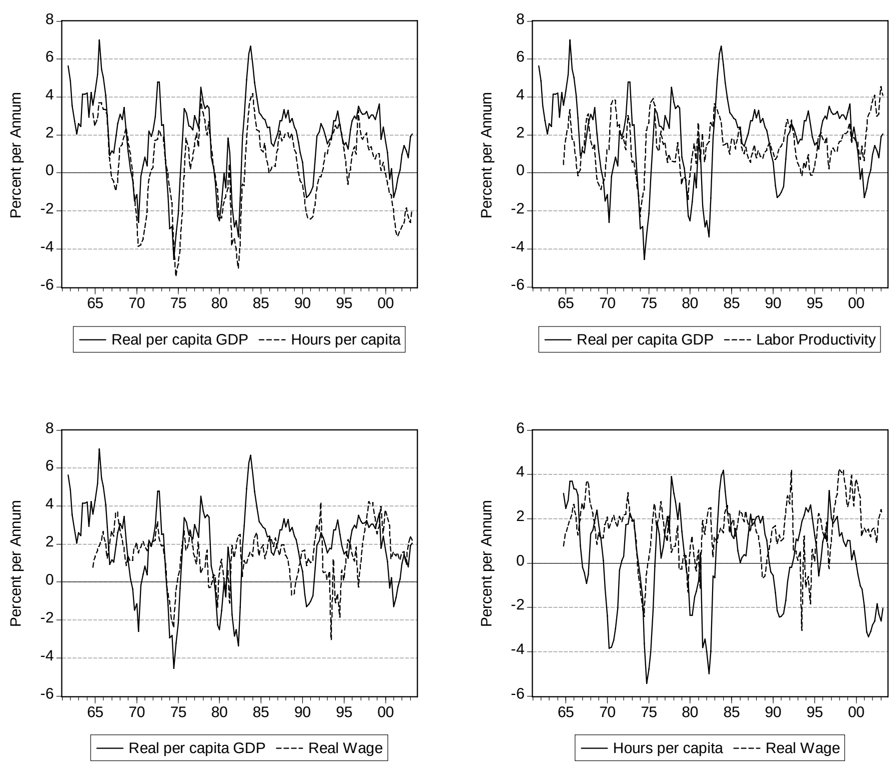
- **Exercise 2.5.** Using a diagram similar to Figure 2.5, demonstrate what effect productivity shocks have on the equilibrium in the labor market.

From Figures 2.5 and 2.7, we see that a positive productivity shock (e.g., an increase in z from z_M to z_H) has the effect of increasing the demand for labor (since labor is now more productive). The increase in labor demand puts upward pressure on the real wage, leading households to substitute time away from home production and into the labor market. Real GDP rises for two reasons: (1) employment is higher; and (2) labor is more productive. Economic welfare also increases. The same results, but in reverse, apply when the economy is hit by a negative productivity shock.

The main prediction of the neoclassical model is that cyclical fluctuations in productivity should naturally result in a business cycle. In other words, one cannot simply conclude (say from Figure 1.4) that the business cycle is necessarily the result of a malfunctioning market system. According to this interpretation, the business cycle reflects the market's *optimal reaction* to exogenous changes in productivity.

One way to 'test' the real business cycle interpretation is to see whether the model's predictions for other economic aggregates are broadly consistent with observation. In particular, note that the model developed above (as with more elaborate versions of the model) predicts that employment (or hours worked), labor productivity, and the real wage should all basically move in the same direction as output over the cycle (i.e., these variables should display *procyclical* behavior). Figure 2.8 displays the cyclical relationships between these variables in the U.S. economy.

FIGURE 2.8
The U.S. Business Cycle:
Growth Rates in Selected Aggregates
(5 quarter moving averages)



From Figure 2.8, we see that hours worked do tend to move in the same direction as output over the cycle, which is consistent with our model. However, the correlation between output and hours is not perfect.[5] Both labor productivity and the real wage are also procyclical, but much less than what our model predicts. But the real problem with our theory appears to be with the behavior of hours worked vis-à-vis the real wage. In this data, at least, there appears little relation between wage movements and hours over the cycle.

2.3.1 The Wage Composition Bias

While the behavior of hours and wages may be taken as a rejection of our simple model, it does not necessarily constitute a rejection of the real business cycle hypothesis. In particular, it is possible to construct more elaborate (i.e., realistic) versions of the neoclassical model that are consistent with the data.

One such extension involves modeling the fact that workers differ in skill (in our version of the model, we assumed that all workers were identical). As it turns out, most of the cyclical fluctuations in hours takes the form of lower-skilled workers moving into and out of employment. Furthermore, the market return to factory work, construction, and other lower skilled trades appears to be much more volatile than, say, the market return enjoyed by doctors and lawyers. To the extent that this is true, then the average real wage (computed as the total wage bill divided by hours worked) is likely to suffer from a *composition bias*.[6]

The composition bias refers to the fact that the composition of skills in the workforce changes over the business cycle. During a cyclical downturn, for example, since lower-skilled workers are more prone to losing their jobs, the average quality of the workforce increases (so that the measured real wage will be higher than if the average quality of the workforce remained unchanged). To see how this might work, consider the following example. Suppose that the economy consists of workers: a fraction $(1-\theta)$ are high-skilled and a fraction θ are low-skilled. High-skilled workers earn a stable wage (since their productivity is relatively stable) equal to w_H. Suppose, however, that the productivity of low-skilled workers fluctuates so that their real wages fluctuate over the cycle; i.e., $w_L(z_H) > w_L(z_L)$. All workers can either work full-time or not (i.e., time is indivisible). The return to home production v is the same for all workers. Assume that:

$$w_H > w_L(z_H) > v > w_L(z_L).$$

In this simple model, when productivity is high (z_H), all individuals will choose to work and when productivity is low (z_L), only high-skilled individuals

[5] In particular, note that during the most recent recovery, output has grown while hours worked have continued to decline. This pattern is also evident (but to a lesser extent) during the recovery of the early 1990s. In the popular press, this type of behavior has come to be called a 'jobless recovery,' and is the subject of much current policy debate. We will return to this issue shortly.

[6] See: Solon, Barsky and Parker (1994).

will work. Thus, the composition of skills in the workforce changes with z, since lower-skilled individuals do not work when productivity is low. Consequently, the 'average' wage among employed workers in this model will be *countercyclical* (i.e., move in the opposite direction of output). That is, when productivity is low, the average real wage is given by w_H. But when productivity is high, the average real wage is given by $(1-\theta)w_H + \theta w_L(z_H) < w_H$. Thus, the fact that the measured 'average' real wage is not strongly procyclical may be more of a measurement problem than a problem with the theory.[7]

2.4 Policy Implications

To the extent that real economies are subjected to exogenous productivity shocks, our model predicts that the economy will display business cycle behavior. The resulting cycles of economic activity induce changes in the *economic welfare* of individuals as measured by the level of utility (notice how the indifference curves in Figure 2.7 move up and down with the cycle).

The business cycle is often viewed as a 'bad' thing. This is certainly true in our model; in particular, our model people would be happier (enjoy higher lifetime utility) in the absence of fluctuations around trend levels. The business cycle is also often viewed as evidence of a 'malfunctioning' market economy; a view that invariably calls for some sort of government intervention to 'fix' the problem. This interpretation and policy advice, however, is not consistent with our model.

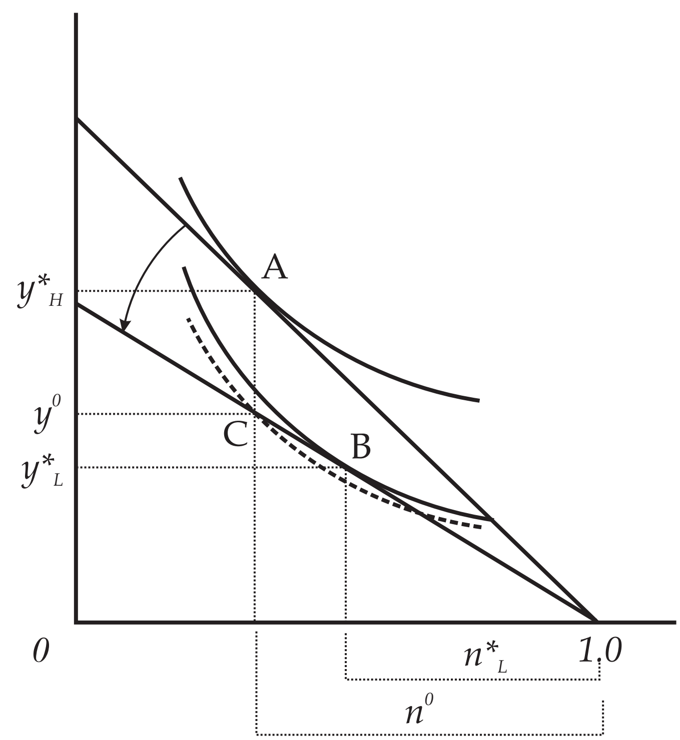
In our model, the equilibrium allocation is in fact *efficient* (in the sense of being *Pareto optimal*). In other words, markets work to allocate resources over the cycle in precisely the 'correct' way. Individuals may not like the fact that productivity fluctuates around trend, *but given that it does*, they respond optimally to the changes in their environment brought about by productivity shocks. The implications for policy are rather stark: the government should do nothing. While it may be in the power for the government to stabilize economic aggregates, any attempt to do so would only make the 'average' person worse off.

To demonstrate this point formally, let us consider the following example. Consider an economy that is initially at the equilibrium point A in Figure 2.11. Now, suppose that productivity falls for some exogenous reason. In the absence of any government intervention, the economy moves into recession (point B in Figure 2.11), as both output and employment decline. Now, suppose that the government has the power to stabilize employment (for example, by subsidizing wage costs for firms). The government does not, however, have the power to prevent the decline in productivity. In this case, the equilibrium allocation would be given by an allocation like C in Figure 2.11. Note that this government

[7] For example, when one corrects for the composition bias, Liu (2003) finds that the return to work is strongly procyclical.

intervention has the effect of stabilizing employment and reducing the decline in GDP. While this may, on the surface, sound like a good thing, note that this government stabilization policy *reduces economic welfare*.

FIGURE 2.11
Government Stabilization Policy



To understand the intuition behind this policy advice, consider the following example. Imagine that our country is hit by an unusually harsh winter (or perhaps a breakdown in the power grid, as happened recently in eastern North America). Among other things, extremely cold weather reduces the productivity of labor in many sectors of the economy. Many firms (like those in the construction sector) reduce their demand for labor, or what amounts to the same thing, reduce the amount they would be willing to pay for less productive labor. Individuals respond to the temporary decline in productivity by reallocating time away from the market sector into the home sector (so that employment and output falls). In principle, the government could step in to 'stabilize' the decline in employment. One way they could do this is by offering wage subsidies (financed by some type of tax). Another more direct way would be to pass legislation forcing individuals to work harder. But what would be the point of such an intervention? The fundamental 'problem' here is the decline in productivity; and there is nothing that the government can do to prevent winter from coming.

2.5 Uncertainty and Rational Expectations

In the neoclassical model, expectations are viewed as passively adjusting themselves to changing economic fundamentals. To understand what this means, let us modify the model above to incorporate an element of uncertainty over the level of productivity z. Imagine that z can take on one of two values: $z_H > z_L$. Let $\pi(z, s)$ denote the probability of z occurring, conditional on receiving some information s (a signal that is correlated with productivity).

The way the economy works is as follows. At the beginning of each period, individuals receive a signal s that is either 'good' or 'bad;' i.e., $s \in \{g, b\}$. If people observe the good signal, then it is known that z_H is more likely than z_L. On the other hand, if people observe the 'bad' signal, then it is known that z_L is more likely than z_H. In other words,

$$\begin{aligned} \pi(z_H, g) &> \pi(z_L, g); \\ \pi(z_H, b) &< \pi(z_L, b). \end{aligned} \tag{2.6}$$

After observing the signal, assume that workers and firms must *commit* to a level of employment (before actually knowing the true level of productivity that occurs). This assumption captures the idea that some investments must be undertaken without knowing for sure what the actual return will be.

A key assumption made in the neoclassical model is that individuals understand the 'fundamentals' governing the random productivity parameter; i.e., they 'know' the function π, as well as the variables (z, s). These fundamentals are viewed as being determined by God or nature. In other words, for better or worse, this is just the way things are; i.e., the world is an uncertain place and people must somehow cope with this fact of life.

So how might individuals cope with such uncertainty? It seems reasonable to suppose that they form *expectations* over z, given whatever information they have available. Since individuals (in our model) are aware of the underlying fundamentals of the economy, they can form 'rational' expectations. Let $z^e(s)$ denote the expected value for z conditional on having the information s. Then it is easy to calculate:

$$\begin{aligned} z^e(g) &= \pi(z_H, g)z_H + \pi(z_L, g)z_L; \\ z^e(b) &= \pi(z_H, b)z_H + \pi(z_L, b)z_L. \end{aligned} \tag{2.7}$$

Given the probability structure described in (2.6), it is clear that $z^e(g) > z^e(b)$. In other words, if people observe the 'good' signal, they are 'optimistic' that productivity is likely to be high. Conversely, if people receive the 'bad' signal, they are 'pessimistic' and believe that productivity is likely to be low. As information varies over time (i.e., as good and bad signals are observed), it will appear as if the 'mood' or 'confidence' of individuals varies over time as well. These apparent mood swings, however, have nothing to with psychology; i.e., they reflect entirely rational changes in expectations that vary as the result of changing fundamentals (information).

In our model, firms always earn (in equilibrium) zero profits–both in an expected and actual (before and after uncertainty is resolved). What this implies is that the expected wage must be given by $w^e(s) = z^e(s)$ and the actual wage must be given by $w^*(z) = z$. By assumption, however, employment decisions must be based on the expected wage. Since actual employment in our model is determined entirely by labor supply, household decision-making determines the level of employment and the (expected) level of consumer demand; i.e.,[8]

$$\begin{aligned} MRS(c^e(s), 1 - n^*(s)) &= w^e(s); \\ c^e(s) &= w^e(s)n^*(s). \end{aligned}$$

In this model, observing a good signal will result in an employment boom (since the expected return to labor is high). Conversely, observing a bad signal will result in an employment bust (since the expected return to labor is low). The economics here are exactly the same as what has been described earlier. The only difference here is that *actual* GDP (consumption) will in general differ from *expected* GDP (consumption). That is, actual GDP is given by $y^*(z, s) = zn^*(s) = c^*(z, s)$.

Thus, it appears that individuals will in general make 'mistakes' in the sense of regretting past (employment) decisions. Note, however, there is nothing 'irrational' about making such mistakes. At the time decisions were to be made, actions were made on the basis of the best possible information available. If this is true, then our conclusion about the role of government stabilization policies continues to hold.

- **Exercise 2.6**. Use a diagram to depict a situation where $s = g$ but where productivity ends up being low (i.e., $z = z_L$).

2.6 Animal Spirits

The neoclassical view that expectations passively adjust to changing economic fundamentals is not one that is shared by all economists and policymakers. In some circles, there is a strong view that expectations appear to be driven by psychological factors and not by changes in economic fundamentals. According

[8] Instructors: For this mathematical characterization to hold exactly, we must assume that certainty equivalence holds (e.g., assume that preferences are quadratic).

to this view, exogenous (i.e., unexplained) changes in expectations themselves constitute an important source of 'shocks' for an economy. These 'expectation shocks' are called *animal spirits*–a colorful phrase coined by Keynes (1936, pp. 161–162):

> "Even apart from the instability due to speculation, there is the instability due to the characteristic of human nature that a large proportion of our positive activities depend on spontaneous optimism rather than mathematical expectations, whether moral or hedonistic or economic. Most, probably, of our decisions to do something positive, the full consequences of which will be drawn out over many days to come, can only be taken as the result of animal spirits - a spontaneous urge to action rather than inaction, and not as the outcome of a weighted average of quantitative benefits multiplied by quantitative probabilities."

As with much of Keynes' writing, the ideas expressed here are simultaneously thought-provoking and ambiguous. This ambiguity is evident in the way scholars have interpreted the notion of animal spirits. One interpretation is that animal spirits constitute 'irrational' (psychological) fluctuations in 'mood' that are largely independent of economic fundamentals.[9] Another interpretation is that animal spirits are driven by psychological factors, but are nevertheless consistent with individual (not necessarily social) rationality to the extent that exogenous shifts in expectations become *self-fulfilling prophesies* (i.e., the economic fundamentals themselves can depend on expectations).[10] In what follows, I try to formalize each of these views.

2.6.1 Irrational Expectations

Let us reconsider the model developed in Section 2.5. One way to model 'irrational' expectations is as follows. Assume, for simplicity, that the true level of productivity remains fixed over time at some level z. In this world then, economic fundamentals remain constant. However, suppose that for some unexplained reason (e.g., 'psychological' factors), individuals form expectations according to (2.7). These expectations are not consistent with the underlying reality of the economy; i.e., they are irrational given the fact that z remains constant over time. Nevertheless, individuals (for some unexplained reason) act on signals that bear no relationship to economic fundamentals (these signals are useless for forecasting productivity).

[9] This interpretation is popular in policy circles. Consider, for example, when Federal Reserve Chairman Alan Greenspan warned us of 'irrational exuberance' in the stock market.

[10] Keynes (1936, pg. 246) himself can be interpreted as adopting this view when he remarks that an economy "seems capable of remaining in a chronic condition of sub-normal activity for a considerable period without any marked tendency either towards recovery or towards complete collapse."

If the world did operate in the way just described, then there would be a business cycle even in the absence of any changes in economic fundamentals. When $s = g$, individuals become 'optimistic' (in fact, they become overly optimistic). This optimism leads to a boom in employment and output. Likewise, when $s = b$, individuals become 'pessimistic' (in fact, they become overly pessimistic). This pessimism leads to a bust in employment and output. These fluctuations in output and employment occur despite the fact that z remains constant.

- **Exercise 2.7**. Use a diagram to depict the economy's general equilibrium when expectations are rational (i.e., the neoclassical view). Now depict a situation in which individuals irrationally act on a signal $s = g$ (revising upward their forecast of productivity). Are individuals made better off by the resulting boom in output and employment? Explain.

Needless to say, if one was inclined to adopt this view of the world, the policy implications are potentially quite different from the neoclassical view. In this world, the business cycle is clearly a 'bad' thing, since it is entirely the product of overly optimistic and overly pessimistic fluctuations in private sector expectations. If the government could design an effective stabilization policy, then such a policy is likely to improve economic welfare. Keep in mind, however, that to design such a policy one must assume that the government in some way has better information concerning the 'true' state of the economy than the private sector. Whether this might be true or not is an empirical question (and one that is not fully resolved).

2.6.2 Self-Fulfilling Prophesies

A self-fulfilling prophesy is a situation where an exogenous change in expectations can alter economic reality in a way that is consistent with the change in expectations. Some of us may have had experiences that fit this description. For example, you wake up on the morning before an exam and (for some unexplained reason) expect to fail the exam regardless of how hard you study. With this expectation in place, it makes no sense to waste time studying (you may as well go to the bar and at least enjoy the company of friends). Of course, the next day you write the exam and fail, confirming your initial expectation (and rationalizing your choice of visiting the bar instead of studying). But suppose instead that you woke up that fateful morning and (for some unexplained reason) expect to pass the exam with a last-minute cram session. The next day you write the exam and pass, confirming your initial expectation (and rationalizing your decision to study rather than drinking).

According to this story, what actually ends up happening (pass or fail) depends critically on what sort of 'mood' you wake up with in the morning before the exam. Your mood is uncontrollable. But given your mood, you can form expectations rationally and act accordingly. Your actions and outcomes can be consist with your initial expectations.

This example is probably not the best one since it relates one's expectation only to oneself. But the same idea can apply to how one's expectations are formed in relation to the behavior of others. Suppose, for example, that the existing technology is such that the return to your own labor is high when everyone else is working hard. Conversely, the return to your labor is low when everyone else is slacking off. Imagine that you must make a decision of how hard to work based on an expectation of how hard everyone else is going to work. If I expect everyone else to work hard, it makes sense for me to do so as well. On the other hand, if I expect everyone else to slack off, then it makes sense for me to do so as well. This is an example of what is called a *strategic complementarity* (Cooper, 1999). A strategic complementarity occurs when the payoff to any action I take depends positively on similar actions taken by everyone else.

If the economy is indeed characterized by strategic complementarities, then the possibility of *coordination failure* exists. To illustrate the idea of coordination failure, imagine that everyone wakes up one morning and (for some unexplained reason) expect everyone else to work hard. Then the equilibrium outcome becomes a self-fulfilling prophesy, since at the individual level it makes sense for everyone to work hard under these expectations. Of course, the converse holds true if everyone (for some unexplained reason) wakes up in the morning expecting everyone else to slack off. In this latter scenario, expectations are coordinated on a 'bad' equilibrium outcome (which is what people mean by coordination failure).

This basic idea can be modeled formally as follows. Imagine that the economy's production technology takes the following form:

$$y = \begin{cases} z_H n & \text{if} \quad n \geq n_C; \\ z_L n & \text{if} \quad n < n_C; \end{cases}$$

where n_C denotes some 'critical' level of *aggregate* employment and n represents the actual level of *aggregate* employment. This technology exhibits a form of *increasing returns to scale*. In particular, if aggregate employment is low (in the sense of being below n_C), then productivity is low as well. Conversely, if aggregate employment is high (in the sense of being at least as large as n_C), then productivity is high as well. Hence, the level of productivity depends critically on whether employment is high or low in this economy.

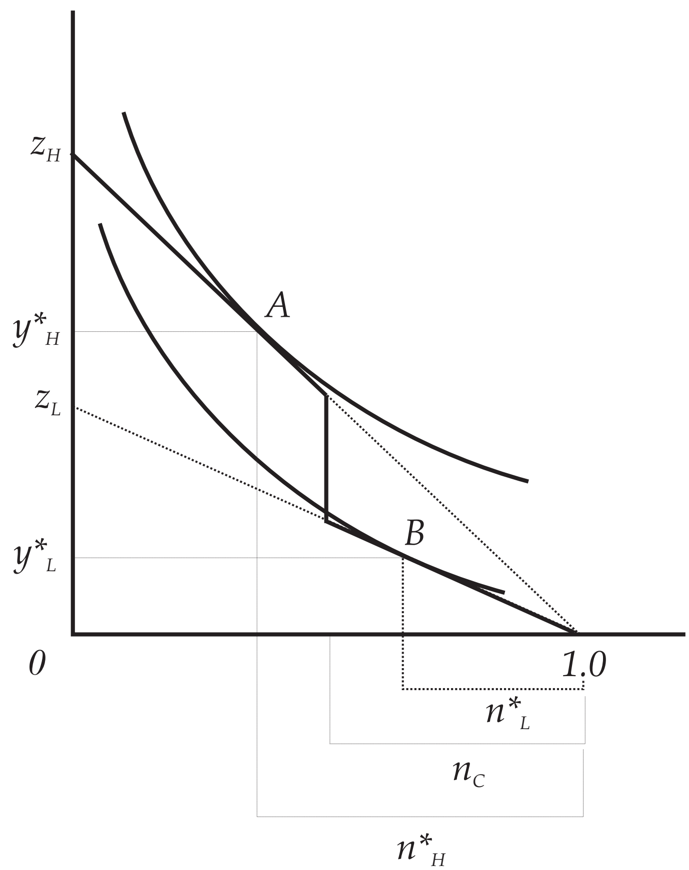
As before, the equilibrium real wage in this economy will equal the marginal product of labor (z). Imagine that households wake up in the morning expecting some z. Then their optimal consumption-labor choices must satisfy:

$$MRS(c^*(z), 1 - n^*(z)) = z;$$
$$c^*(z) = zn^*(z).$$

In equilibrium, the *aggregate* (average) level of employment must correspond to each household's *individual* employment choice. Assume that $n^*(z_L) < n_C <$

$n^*(z_H)$. These two possible outcomes are displayed in Figure 2.12 as points A and B.

FIGURE 2.12
Multiple Equilibria



Notice that both points A and B are consistent with a rational expectations general equilibrium. In the neoclassical model studied earlier, the general equilibrium was unique (owing to the constant returns to scale in the production technology). In the model studied here, however, there are *multiple equilibria*. Each equilibrium is generated by a self-fulfilling prophesy. Individuals are clearly better off in equilibrium A. However, if individuals coordinate their expectations on B, then this outcome too is a possibility. Whether outcome A or B occurs depends on an arbitrary set of initial expectations. Exogenous changes in these initial expectations may cause the economy to fluctuate between points A and B over time. On the other hand, if beliefs become 'stuck' at point B, the economy may experience a prolonged self-fulfilling 'depression.'[11]

- **Exercise 2.8**. Consider Figure 2.12. Suppose you expect a low level of aggregate employment and output. Suppose, however, that you decided to 'work hard;' i.e., choose $n = n_H$. Draw an indifference curve that shows the level of consumption (and utility) you can expect to enjoy and explain why it is not rational to make such a choice. Suppose instead that you expect a high level of aggregate employment and output. Further, suppose that you decide to 'slack off;' i.e., choose $n = n_L$. Again, draw an indifference that shows the level of consumption (and utility) you can expect to enjoy and explain why it is not rational for you to make such a choice.

As with the earlier interpretation of animal spirits, this view of the world suggests a potential role for government policy. Suppose, for example, the economy appears to be stuck at a 'bad' equilibrium (point B). In this situation, the 'demand' for output appears to be weak and the level of employment is low. Such a scenario may rationalize a large fiscal expenditure on the part of the government. The government might do this by placing orders for output in the market, or by hiring individuals directly to work in government agencies that produce output. Either way, the demand for labor can be increased beyond the threshold n_C. By doing so, individual expectations must rationally move from point B to point C.

2.7 Summary

The neoclassical view of the business cycle is that fluctuations are induced by the actions of rational agents in response to exogenous changes brought about the natural process of technological development. Economic expansions are periods where the expected return to market activity is high. Recessions are periods where the expected return to market activity is low. Expectations may fluctuate even in the absence of any immediate changes in underlying productivity (as people may receive information leading them to revise their estimates on the future returns to current investments). While expectations play a critical role in the neoclassical view, private sector expectations only shift in response to perceived changes in the underlying economic fundamentals in an economy. Often, expectations are wrong; i.e., things may turn out better or worse than expected. Nevertheless, expectations are viewed as being 'rational' in the sense of being formed in a way that is consistent with the underlying probability structure governing the realizations of random events. In this view of the world, government stabilization policies are likely to do more harm than good.

An alternative view of the cycle suggests that expectations may have a 'life of their own' in the sense that they may shift for 'psychological' reasons that

[11] From the quote from Keynes in footnote 9, this scenario appears to describe his interpretation of the Great Depression.

may not be related in any way to changing economic fundamentals. One version of this view asserts that expectations are 'irrational' in the sense of bearing no relationship whatsoever to fundamentals. Another version of this view asserts that exogenous changes in expectations may induce self-fulfilling prophesies. In either of these cases, there is a clear potential role for government stabilization policies. Note that this conclusion can hold even if markets are assumed to 'clear' in the neoclassical sense (as is the case with all the models developed in this chapter). But whether the government can, as a practical matter, identify an 'irrational' expectation or the existence of a self-fulfilling prophesy remains open to question.

2.8 Problems

1. You are asked to write a brief newspaper column explaining the real business cycle interpretation of the business cycle. Obviously, you cannot use mathematical notation or make any use of diagrams. Furthermore, your audience will generally not be familiar with economic jargon. Limit yourself to 800 words or less.

2. Why is it important for any economic model to have the preferences of individuals stated explicitly?

3. Does the fact that the expectations of private agents may differ from actual outcomes necessarily imply a role for government policy? Explain.

4. Which of the two versions of the 'animal spirits' hypothesis described in the text do you find more plausible? Explain why you feel this way (what sort of evidence do you have to support your view)?

5. Imagine that a technological advance appears that favors some sectors of the economy while hurting others. Should the government attempt to stabilize the level of employment in declining sectors? How would such a policy likely to affect growth in an economy? What other sort of policy might the government undertake to facilitate the movement of factors from declining industries to expanding sectors of the economy?

2.9 References

1. Cooper, Russell (1999). *Coordination Games: Complementarities and Macroeconomics*, New York: Cambridge University Press.

2. Keynes, John M. (1936). *The General Theory of Employment*, Interest and Money, MacMillan, Cambridge University Press.

3. Kydland, Finn and Edward C. Prescott (1982). "Time to Build and Aggregate Fluctuations," *Econometrica*, 50(6): 1145–1170.

4. Liu, Haoming (2003). "A Cross-Country Comparison of the Cyclicality of Real Wages, *Canadian Journal of Economics*, 36(4): 923–948.

5. Long, John B. and Charles Plosser (1983). "Real Business Cycles," *Journal of Political Economy*, 91(1): 39–69.

6. Schumpeter, Joseph A. (1942). *Capitalism, Socialism, and Democracy*, Reprinted by Harper-Collins (1975).

7. Solon, Gary, Barsky, Robert and Jonathan A. Parker (1994). "Measuring the Cyclicality of Real Wages: How Important is the Composition Bias?" *Quarterly Journal of Economics*, 109(1): 1–25.

2.A A Model with Capital and Labor

The model of time allocation developed in this chapter assumed that labor is the only factor of production. If there are constant returns to scale in production, then this motivates the existence of a production function that takes the form $y = zn$.

The model is easily extended to include two factors of production, which we can think of as capital (k) and labor (n). In this case, the production function can be written in a general form as $y = zF(k, n)$. A specific functional form for F is the *Cobb-Douglas* specification: $F(k, n) = k^{1-\theta}n^{\theta}$, where $0 \leq \theta \leq 1$ is a parameter that indexes the relative importance of labor in production. Our earlier model is just the special case in which $\theta = 1$.

As this is a static model, assume that the amount of capital is fixed in supply; for example, suppose that $k = 1$. The production function can then be written as $y = zn^{\theta}$. What this tells us is that output is an increasing function of labor (as before). However, as employment expands, output does not expand in a linear manner (as it did before). In particular, output expands at a declining rate. The reason for this is because as more labor works with a given amount of capital, the average (and marginal) product of labor declines.

For this production function, the marginal product of labor is given by:

$$MPL(n, z) = \theta z n^{\theta-1}.$$

The *marginal product of labor* tells us the extra output that can be produced with one additional unit of labor. On a diagram, the MPL can be depicted as the slope of the production function. Observe that when $\theta = 1$, we have $MPL(n, z) = z$ (i.e., the MPL is a constant). When $\theta < 1$, then the MPL declines (the slope of the production function becomes flatter) as n increases. As well, note that for a given level of n, an increase in z implies an increase in the MPL.

The choice problem facing a typical firm is given by:

Choose (n) in order to maximize $d = zn^{\theta} - wn$.

The solution to this choice problem is a desired labor input (labor demand) function n^D, which happens to satisfy the following condition:

$$MPL(n^D, z) = w.$$

As an exercise, you should try to solve for the labor demand function n^D and show that it is decreasing in w and increasing in z. Once we know n^D, we can easily calculate the supply of output $y^S = z(n^D)^{\theta}$ and the planned dividend payment $d = y^S - wn^D$. Note that unlike before, firms here will generally earn a positive profit $d > 0$, which reflects the return to the capital used in production.

The individual's choice problem remains as before, except that now households have two sources of income (wage and dividend income). Given some arbitrary market wage w, the solution satisfies:

$$MRS(c^D, l^D) = w;$$
$$c^D = w(1 - l^D) + d.$$

Once we know l^D, we can infer the labor supply function from the time-constraint: $n^S = 1 - l^D$.

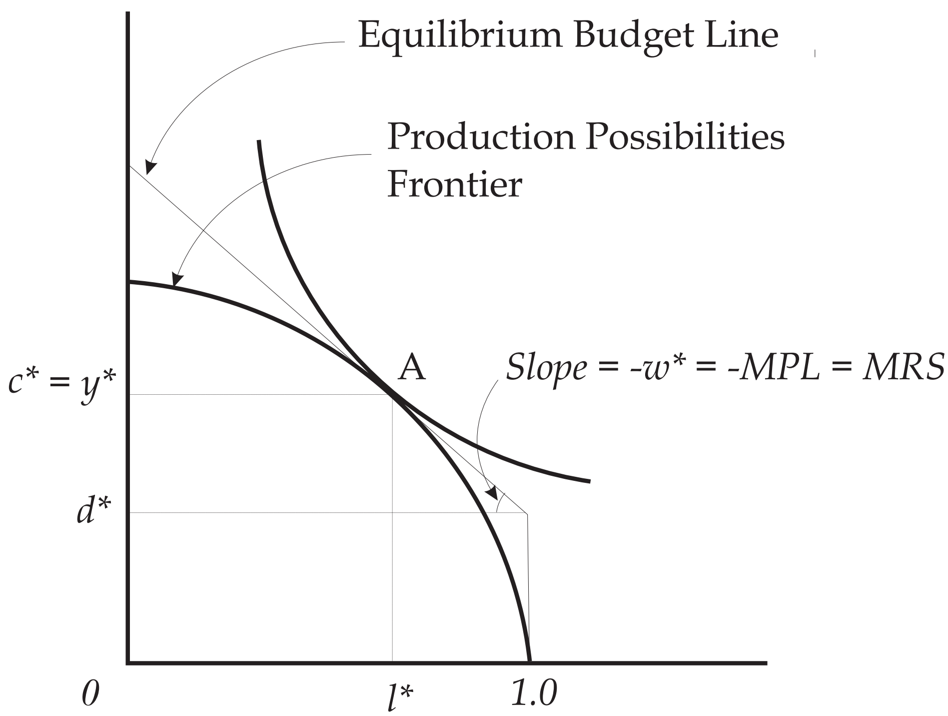
Since the market wage w has been arbitrarily chosen at this stage, it is generally not a market-clearing wage. The next step then is to impose the *market-clearing condition*:

$$n^D = n^S.$$

The assumption here is that the real wage adjusts in order to ensure that the market clears (again, remember that there is only one market in this model). The model's general equilibrium is depicted as Point A in Figure 2.12.

The general equilibrium of this model economy differs from the one in the text in only two minor ways. First, the equilibrium real wage is equal to $w^* = z(n^*)^{\theta-1}$ (it now varies with the equilibrium level of employment). When $\theta = 1$, we once again have $w^* = z$. Second, firms actually earn 'profit' (generate dividends) in this model. This profit represents the return to the capital that is used in production. Since households own the equity in the business sector, they are also the ultimate owners of capital. The dividend payment reflects this ownership in the capital stock.

FIGURE 2.12
General Equilibrium



2.B Schumpeter's Process of Creative Destruction

Capitalism's Forces of Creative Destruction Unleash Opportunities for Investors

by Frederick B. Taylor, Vice Chairman and Chief Investment Officer

Updated: 4/9/02

Writing about capitalism more than 50 years ago, economist Joseph Schumpeter coined the phrase "creative destruction" to describe the process by which a free-market economy is constantly evolving, as new and better ways of doing business are introduced and the old and outmoded fall by the wayside. Creative destruction "revolutionizes the economic structure from within," said Schumpeter, "destroying the old one, (and) creating a new one." He argued that this dynamic process was central to capitalist system's ability to maximize output and total wealth creation over time.

Creative destruction is not a steady process. While the forces of creative destruction are always present in a capitalist system, the process often occurs in intense bursts - "discrete rushes," as Schumpeter termed them, "which are separated from each other by spans of comparative quiet." Historically, we experienced a period of intense creative destruction in the late 19th and early 20th centuries, and we are now living in the midst of another. At the turn of the last century, the industrial revolution had transformed agricultural economies and was radically changing most people's lives. New industrial powers—the United States and Germany—were developing to challenge the established economies of England and France. Today, the ongoing revolution in technology and communications is creating new industries and transforming or eliminating old ones. New economies are emerging that provide new markets as well as new sources of competition. Corporate restructuring and privatization of government-owned enterprises are widespread.

Usually, one can only see what is being destroyed by this process; it is much more difficult to understand what is replacing it. For example, it was easy to see carriage and buggy whip manufacturers going out of business with the advent of the automobile, but much harder to visualize the mega-industry that would emerge to replace them. Similarly, as manufacturing declined in the U.S. during the latter part of the 20th century, the displacement of factory workers was evident long before the creation of jobs in the service sector that reemployed them.

Although the process of creative destruction can create tremendous opportunities for investors, it is often difficult to discern them early on, when the maximum benefit can be gained. However, in our search for value, this is exactly what we strive to do for our clients.

On a global level, those countries where capitalism and the forces of creative destruction are permitted to flourish create more attractive opportunities for investors than less dynamic economies. Of course, the United States continues to be the prime example of relatively unfettered capitalism at work, and the U.S. economy has been better able not only to withstand, but to thrive amid the current forces of creative destruction—hence the sustained strong performance of the U.S. stock market in the 1990s. However, there are other countries that are also benefiting from the forces of creative destruction, most notably in Europe. Great Britain was the first to dismantle the government programs that had hindered its economic growth. As the rest of Europe follows suit, albeit each country at its own speed, creative destruction will be at work restructuring old inefficient economies into more vibrant, growing ones. In particular, Ireland, Italy, Portugal, and Spain, which have less rigid economic systems than France and Germany, are learning to compete effectively in the global economy. The introduction of the euro and the emergence of Pan-Europeanism are also expanding markets for companies able to compete effectively in this new environment, and the "creative destruction" of national currencies has given birth to the Euro. A few Latin American countries, most notably Argentina and Chile, are becoming free-market economies. And in Asia, where we recently witnessed the destruction of the old status quo, opportunities are beginning to emerge for innovative companies.

At the root of the creative destruction process are individual companies. They are the agents of change that develop new products, new technology, new production or distribution methods, new markets and new types of organization that will revolutionize the economy. Companies that change the business model for their industry or develop a new paradigm significantly alter the competitive landscape. As Schumpeter says, "competition which commands a decisive cost or quality advantage... strikes not at the margins of the profits and the outputs of the existing firms but at their foundations and their very lives."

We want to own companies that benefit from creative destruction. Good businesses that are constantly adapting to their changing environment and establish a decided competitive advantage will generate strong returns for their shareholders over a long period of time. We search industries throughout the world to find those companies that are transforming the terms of competition in their field. The forces of creative destruction in capitalism are not only the engine of economic growth; they also generate unparalleled investment opportunities for astute investors.

1. From Joseph A. Schumpeter, *History of Economic Analysis*, edited from a manuscript by Elizabeth Boody Schumpeter (New York: Oxford University Press, 1963).

Chapter 3

Fiscal Policy

3.1 Introduction

In this chapter, we extend the basic neoclassical model to include a government sector that demands some fraction of the economy's output for some purpose that we view as being determined exogenously by political factors. Note that this chapter does not constitute a theory of government. Rather, the theory developed here is designed to explain how the economy reacts to any given (exogenous) change in a government's desire to tax and spend. We continue to work with a static model, so that the government must balance its budget on a period-by-period basis (i.e., the government may not run surpluses or deficits). The issue of deficit-finance will be addressed in a later chapter when we have the tools to investigate dynamic decision-making.

3.2 Government Purchases

Assume that the government sector 'demands' g units of output, where g is exogenous. There are two ways of viewing the production of output destined for the government sector. The first way is to suppose that all output y is produced by the private sector and that the government sector purchases the output it desires g from the private sector. The second way is to suppose that the government produces the output it needs by employing workers (public sector workers). If the government has access to the same technology as the private sector, then either approach will yield identical results.

Before proceeding further, we need to ask the question: What is g used for? In reality, government purchases are directed toward a wide variety of uses, including: bureaucratic services, in-kind transfers (e.g., school lunch programs, health-care services), military expenditures, and outright theft (politicians lining their own pockets). For some types of expenditures (e.g., those expenditures that are distributed free of charge to the general population), it would make sense to think of households viewing g as a close substitute for goods and services that they might otherwise purchase from the private sector. In this case, one could model preferences as:

$$u(c+g,l). \tag{3.1}$$

In this case, output that is purchased from the private sector and output that is supplied by the government are viewed as perfect substitutes by the household (e.g., as is likely the case with school lunch programs). If the school lunches supplied by the government do not exactly correspond to the lunches that parents would pack for their kids on their own, then might instead specify $u(c+\lambda g,l)$, with $\lambda \leq 1$. According to this specification, one unit of government supplied output is equivalent to λ units of privately purchased output as far as the household is concerned.

Alternatively, government expenditures may be allocated toward uses that do not have very close substitutes in the way of market goods; e.g., military expenditures or a national space program. In this case, one might more reasonably model preferences as:

$$u(c,l)+\lambda v(g), \tag{3.2}$$

where $\lambda \geq 0$ and v is an increasing and concave function. According to this specification, households may value a national space program not for material reasons but for 'psychic' reasons (e.g., national pride). On the other hand, if households do not value such expenditures at all (e.g., if g is used to build a king's castle, or used to finance an unpopular war), then one could specify $\lambda = 0$.

In what follows, we will adopt the latter specification of preferences (3.2) for the case in which $\lambda = 0$. As it turns out, the model's predictions for economic behavior will continue to hold even for the case in which $\lambda > 0$ (the only thing that would change is the prediction for economic welfare). On the other hand, the model's predictions concerning behavior will generally differ if we adopt the specification (3.1).

3.2.1 Lump-Sum Taxes

Suppose that government spending is financed with a lump-sum tax τ. A *lump-sum tax* is a tax that is placed on individuals that does not vary with the level of their economic activity. For this reason, it is also sometimes called a 'head tax.' Lump-sum taxes are not that common in reality, but for our purposes, it serves as a useful benchmark. The key restriction on government behavior is given by the government budget constraint (GBC), which in this case takes the form:

$$\tau = g. \tag{3.3}$$

As it turns out, in our model, the government fiscal policy will have no effect on the equilibrium wage or profits; i.e., $(w^*, d^*) = (z, 0)$. In general, this result will not hold, but does so here because of the linearity that we have assumed in the production function. But the implications of the fiscal policy on output and employment are not affected by this simplification and so we proceed with this simple specification.

In order to see how individuals are affected by the fiscal policy, we have to restate their choice problem. If individuals must pay a lump-sum tax τ, then their budget constraint is now given by: $c = wn - \tau$. Substituting the time-constraint $n + l = 1$ into the budget constraint allows us to rewrite the budget constraint as: $c = w - wl - \tau$. If the private sector derives no direct utility from g, then the choice problem of individuals is given by:

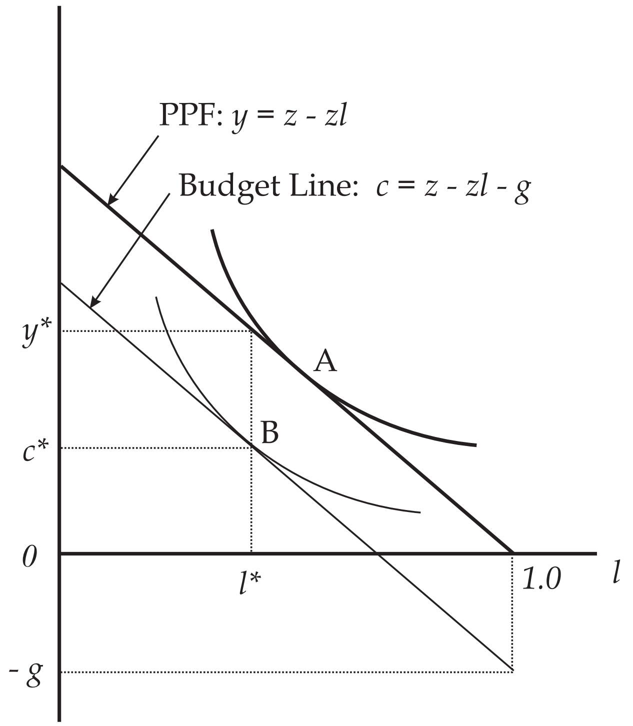
> Choose (c, l) in order to maximize $u(c, l)$
> subject to: $c = w - wl - \tau$ and $0 \leq l \leq 1$.

The solution to this choice problem is a pair of demand functions: $c^D(w, \tau)$ and $l^D(w, \tau)$. Once the demand for leisure is known, one can compute the supply of labor $n^S(w, \tau) = 1 - l^D(w, \tau)$.

Since the choice problem of the business sector is unaffected, we know that the equilibrium wage and dividend payment will equal $(w^*, d^*) = (z, 0)$. As well, the government budget constraint (3.3) will have to be satisfied in equilibrium.

The model we studied in Chapter 2 is just a special case of the model developed here (i.e., in the earlier model, we were assuming $g = 0$). Figure 3.1 displays the general equilibrium of the economy when $g = 0$ (point A). When $g = 0$, the equilibrium budget constraint corresponds to the production possibilities frontier. Suppose now that the government embarks on an 'expansionary' fiscal policy by increasing spending to some positive level $g > 0$. Because the government spending program requires a tax on individuals, their budget constraint moves downward in a parallel manner. If consumption and leisure are *normal goods* (a reasonable assumption), then the new general equilibrium is given by point B in Figure 3.1. Notice that the income-expenditure identity $y = c + g$ is always satisfied in this economy.

FIGURE 3.1
Expansionary Fiscal Policy
Financed with Lump-Sum Tax



The model predicts that an expansionary fiscal policy *financed by a lump-sum tax* induces an increase in output and employment. However, the policy also induces a decline in consumer spending. The basic force at work here is a *pure wealth effect* (there is no substitution effect here because the real wage remains unchanged). In particular, because the after-tax wealth of individuals declines, they demand less consumption and leisure (assuming that both consumption and leisure are normal goods). The decline in the demand for leisure implies that the supply of labor rises. Because firms simply hire all the labor that is supplied in our model, aggregate employment expands, which is what leads to an expansion in real GDP.

Note that while this expansionary fiscal policy causes GDP to rise, it makes our average (model) person worse off (the indifference curve falls to a lower level). There is an important lesson here: be careful not to confuse GDP with economic welfare.

3.2.2 Distortionary Taxation

Now, some of you may be asking whether we needed to go through all the trouble of developing a model to tell us that an increase in g leads to an increase in y. After all, we know from the income-expenditure identity that $y \equiv c + g$. Does it not simply follow from this relationship (which must always hold true) that an increase in g must *necessarily* lead to an increase in y? The answer is *no*.

To demonstrate this claim, let us reconsider the effects of an expansionary fiscal policy when government spending is financed with a *distortionary tax*. A distortionary tax is (as the name suggests) a tax that distorts individual decisions, since the amount of tax that is paid depends on the level of individual economic activity. A primary example of a distortionary tax is an *income tax* (the amount of tax paid depends on how much income is generated). In reality, most taxes are distortionary in nature.

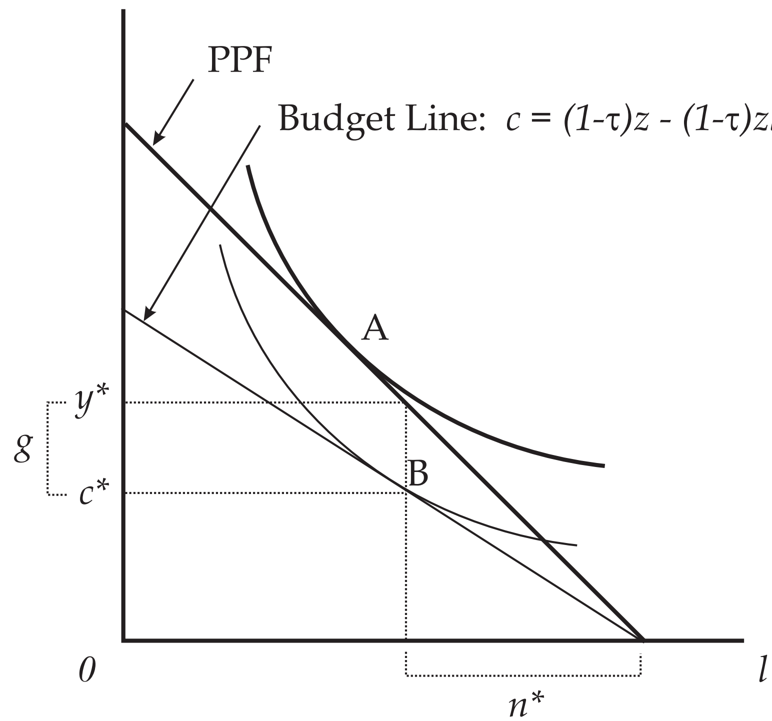
With an income tax, the individual budget constraint becomes: $c = (1 - \tau)wn$, where $0 \leq \tau \leq 1$ now denotes the *income-tax rate*. Substituting the time-constraint $n + l = 1$ into the budget constraint allows us to rewrite the budget constraint as: $c = (1-\tau)w - (1-\tau)wl$. Note that the direct effect of the income tax is to reduce the slope of the individual budget constraint; i.e., the slope is now given by the *after-tax wage rate* $-(1 - \tau)w$. We know that when the slope of the budget constraint changes, there will generally be both a *substitution* and *wealth* effect at work.

Again, the choice problem of the business sector remains unchanged so that $(w^*, d^*) = (z, 0)$. In equilibrium, the government's budget constraint must be satisfied; i.e., the fiscal authority must choose a $\tau = \tau^* > 0$ that satisfies:

$$\tau^* w^* n^* = g. \tag{3.4}$$

Figure 3.2 depicts the original equilibrium when $g = 0$ (point A) and the new equilibrium where $g > 0$ (point B). The position of point B assumes that the substitution effect dominates the wealth effect (which is certainly plausible). Observe now that the effect of an expansionary fiscal policy is to cause a *decline* in both output and employment. The economic intuition of this result is straightforward. The increase in government purchases requires an increase in the income-tax rate, which reduces the after-tax wage rate of individuals. The tax now has two effects. Because individuals are poorer, they reduce their demand for both consumption and leisure, so that labor supply increases (this is the wealth effect). On the other hand, the tax also reduces the price of leisure (increases the price of consumption), so that individuals substitute out of consumption into leisure, thereby reducing labor supply (this is the substitution effect). If the substitution effect on labor supply is stronger than the wealth effect, then the net effect is for employment (and hence output) to fall.

FIGURE 3.2
Expansionary Fiscal Policy
Financed with Income Tax



Observe that it is still true for this economy that $y = c + g$. What has happened here, however, is that the increase in g is more than offset by the resulting decline in c. So while the income-expenditure identity holds true, it is clear that one can not rely simply on this relation to make predictive statements. This is because the income-expenditure identity is not a theory (recall the discussion in Chapter 1).

3.3 Government and Redistribution

In economies like Canada and the United States, government purchases over the last few decades have averaged in the neighborhood of 20% of GDP; see Figure 1.2. Most of these purchases are allocated toward what one might label 'public' goods (output designated for mass consumption), such public parks, public infrastructure (roads, bridges, etc.), public health and education, and national defense. It is interesting to note, however, that these types of government purchases constitute a relatively small fraction of total government spending. In

Canada, for example, total government spending is very close to 50% of GDP (as it is in many European countries). Much, if not most, of this additional spending is in the form of transfer payments to individuals (e.g., unemployment insurance, welfare, pensions, grants, personal and business subsidies, etc.).[1] Since the benefits of such transfers are typically concentrated among specific groups and since the costs are borne by the general public, such transfers necessarily involve some amount of redistribution. A cynic might argue that governments are primarily in the business of robbing Peter to pay Paul.[2] An apologist might argue that government is (or should be) more like Robin Hood (redressing past injustices). In any case, there is no question that redistribution, for better or worse, appears to be a primary function of government. The question we wish to address here is: What are the macroeconomic consequences of policies designed to redistribute income?

In the present context, the government's budget constraint is as follows:

$$\text{Purchases} + \text{Transfers} = \text{Taxes}.$$

The sum (Taxes - Transfers) is referred to as *Net Taxes*. Thus, in our formulation of the government budget constraints (3.3) and (3.4), the expenditure side referred to purchases and the revenue side referred to net taxes. Total government spending here is defined as purchases plus transfers.

To begin, let us recall our benchmark allocation (y^*, l^*), which satisfies $MRS(y^*, l^*) = z$ and $y^* = z(1 - l^*)$. Consider now a government policy that grants each person a lump-sum transfer equal to a units of output. The government finances this transfer with an income tax τ (which we will take as the exogenous policy parameter). For simplicity, assume that all households are identical and that $g = 0$. The household's budget constraint becomes:

$$y = (1 - \tau)z(1 - l) + a, \tag{3.5}$$

with $0 \leq l \leq 1$. The household's optimal choice (y^D, l^D) must lie on this budget line and must satisfy the condition $MRS(y^D, l^D) = (1 - \tau)z$ (in the case that $l^D < 1$).

The government's budget constraint is given by:

$$g + a = \tau z(1 - l), \tag{3.6a}$$

where in this example, $g = 0$. Inserting (3.6a) into (3.5), we see that the equilibrium household's choice (y^0, l^0) must satisfy:

$$y^0 = z(1 - l^0).$$

[1] Note that transfer payments are not counted as part of the GDP since they only serve to redistribute income (and not create additional value).

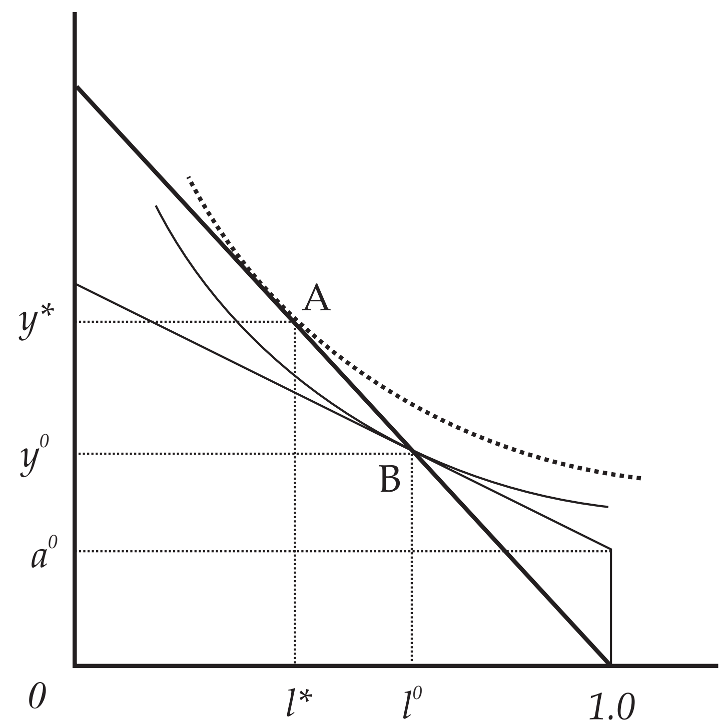
[2] The behavior of some politicians remind me of a radio ad from an establishment in London (Ontario) which proclaimed: "Come to Joe Kool's...where we screw the other guy...and pass the savings on to you!"

In other words, as with the benchmark allocation (y^*, l^*), the household's choice (y^0, l^0) must lie on the PPF. However, the allocation that arises under the redistribution scheme does not correspond to the benchmark allocation; i.e.,

$$MRS(y^*, l^*) = z > (1 - \tau)z = MRS(y^0, l^0).$$

From this condition, we can deduce that $y^0 < y^*$ and $l^0 > l^*$, as in Figure 3.3.

FIGURE 3.3
Lump-Sum Transfer Financed with an Income Tax



Thus, our theory predicts that economies with 'generous' transfer programs (and high average tax rates to finance such programs) should exhibit relatively low levels of real GDP and employment. The intuition is simple. First, the lump-sum transfers reduce labor supply via a pure wealth effect (people do not have to work as hard to acquire output). Second, the income tax distorts the relative price of output and leisure. In particular, since the after-tax wage falls, the relative price of leisure falls so that individuals substitute out of consumption and into leisure. Both of these forces work to reduce GDP and employment.

In the context of our example above that features a 'representative' household, we see that such a government policy also leads to a reduction in welfare.

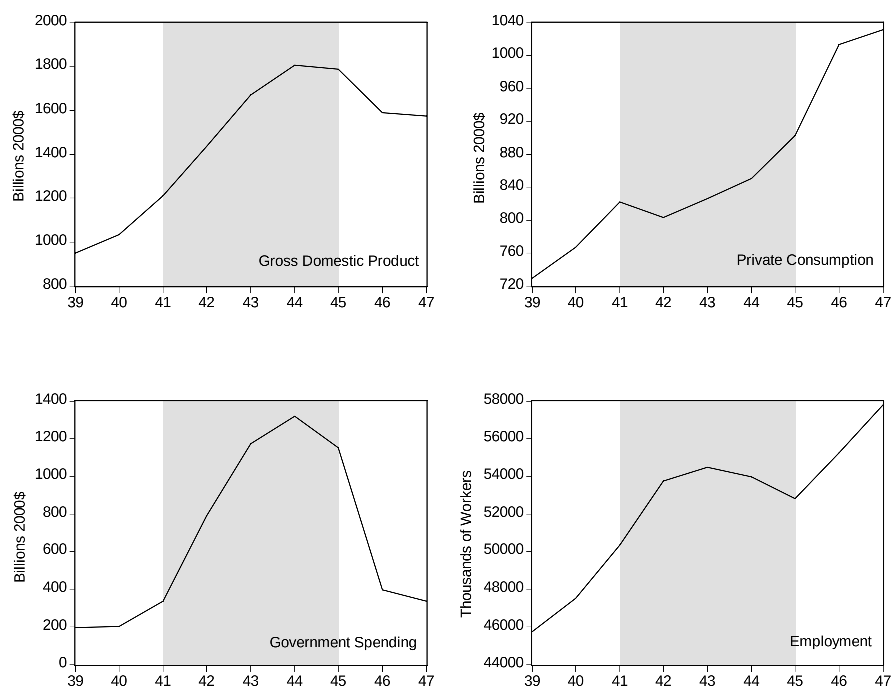
However, in a more realistic model that featured different types of households (e.g., high-skilled versus low-skilled), we would find that some households would benefit while others would lose. The model's predictions concerning the effect on GDP and aggregate employment, however, would remain intact. Note that when households differ, it becomes very difficult for us to make any definitive statement about whether such a redistribution program improves 'social' welfare. The notion of 'social' welfare is ultimately in the eyes of the beholder.

3.4 Problems

1. Suppose that preferences are such that $MRS = c/l$ and that the individual's budget constraint is given by $c = w - wl - \tau$ (a lump sum tax). Derive this person's labor supply function $n^S(w, \tau)$ and explain (provide economic intuition) how it depends on τ.

2. Periods of war often entail huge increases in government military spending financed with government bonds (in the context of our model, you can think of bond-financing as representing a type of lump-sum tax); see Figure 3.2 for the case of the United States during World War II. Are the patterns of economic activity in Figure 3.2 consistent with our theory? Explain. Hint: draw 'trend' lines through the data in the diagrams below.

Figure 3.4
GDP and Employment in the
United States During WWII



3. Suppose that preferences are such that $MRS = (c/l)^{1/2}$ and that the individual faces a budget constraint $c = (1 - \tau)w(1 - l)$. Derive this individual's labor supply function and explain how it depends on τ. Contrast

this result with the result in Question 1. Explain why the results here differ.

4. Does a distortionary tax necessarily imply an adverse effect on labor supply? Derive the labor supply function for preferences given by $MRS = c/l$ and a budget constraint $c = (1-\tau)w(1-l)$. How does an increase in τ affect labor supply here? Explain.

5. Figure 3.3 displays the allocation that would result under a policy that distributes a lump-sum transfer financed by an income tax. The text asserted that the new allocation (B) must lie to the right of the benchmark allocation (A). Prove that this must be the case. Hint: Show that an allocation that lies to the left of (A) must necessarily entail crossing indifference curves.

Chapter 4

Consumption and Saving

4.1 Introduction

Thus far, we have focussed primarily on what one might term *intratemporal* decisions and how such decisions determine the level of GDP and employment at any point in time. An intratemporal decision concerns the problem of allocating resources (like time) across different activities *within* a period. However, many (if not most) decisions have an *intertemporal* aspect to them. An intertemporal decision concerns the problem of allocating resources *across* time. For example, deciding how much to consume today can have implications for how much will be available to consume tomorrow. The decision of how much to invest must be made with a view as to how this current sacrifice is likely to pay off at some future date. If a government runs a deficit today, it must have in mind how the deficit is to be paid off in the future, and so on. Such decisions are inherently *dynamic* in nature. In order to understand how such decisions are made, we need to develop a dynamic model.

In this chapter, we focus on the consumption-savings choice of individuals. Since any act of saving serves to reduce consumption in the present and increase consumption in the future, the key decision involves how to optimally allocate consumption across time. We will study this choice problem within the context of a two-period model. The basic insights to be gleaned from a simple two-period model continue to hold true in a more realistic model that features many periods. In order to focus on the intertemporal aspect of decision-making, we abstract from intratemporal decisions. In particular, the working assumption here is that intratemporal decisions are independent of intertemporal decisions. This assumption is made primarily for simplicity and can be relaxed once the basic ideas presented here are well understood.

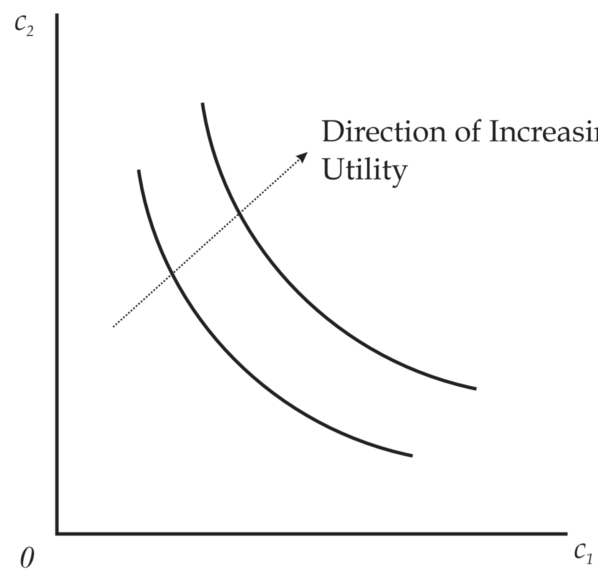
4.2 A Two-Period Endowment Economy

4.2.1 Preferences

Consider a model economy populated by a fixed number of individuals that live for two periods. These individuals have preferences defined over *time-dated* output (in the form of consumer goods and services). Let (c_1, c_2) denote an individual's lifetime consumption profile, where c_1 denotes 'current' consumption and c_2 denotes 'future' consumption. Note that consumption today is *not the same* as consumption tomorrow; *they are treated here as two distinct goods.* The assumption that people have preferences for time-dated consumption simply reflects the plausible notion that people care not only for their material well-being today, but what they expect in terms of their material well-being in the future. In what follows, we assume that there is no uncertainty over how the future evolves.

A lifetime consumption profile (c_1, c_2) can be thought of as a *commodity bundle*. The *commodity space* then is defined to be the space on non-negative commodity bundles and can be represented with a two-dimensional graph. We make the usual assumptions about preferences; i.e., more is preferred to less, transitivity, and convexity. We will also make the reasonable assumption that consumption at different dates are normal goods and that preferences can be represented with a utility function $u(c_1, c_2)$. Figure 4.1 depicts an individual's indifference curves in the commodity space.

FIGURE 4.1
Indifference Curves



4.2.2 Constraints

Individuals are endowed with an earnings stream (y_1, y_2). One can interpret y_j as the real per capita GDP in period $j = 1, 2$. From Chapter 2, we know that the level of GDP is determined by an intratemporal decision concerning the allocation of time to market-sector activities. Because we want to abstract from intratemporal decisions here, let us assume here for simplicity that these decisions are exogenously determined. Given these decisions, the individuals operate *as if* they are faced with an exogenous earnings profile (y_1, y_2). This assumption is reflected in our use of the label: an *endowment* economy.

The assumption that real per capita GDP is exogenous is not important for our purpose here. However, we make one more assumption that does turn out to be important; i.e., that output is non-storable. This is to say that output can not be held in the form inventory. Nor can it take the form of new capital goods, like business fixed investment. For this reason, it is perhaps best to think of output as taking the form of services and perishable goods. We will relax this assumption in a later chapter that discusses capital and investment.

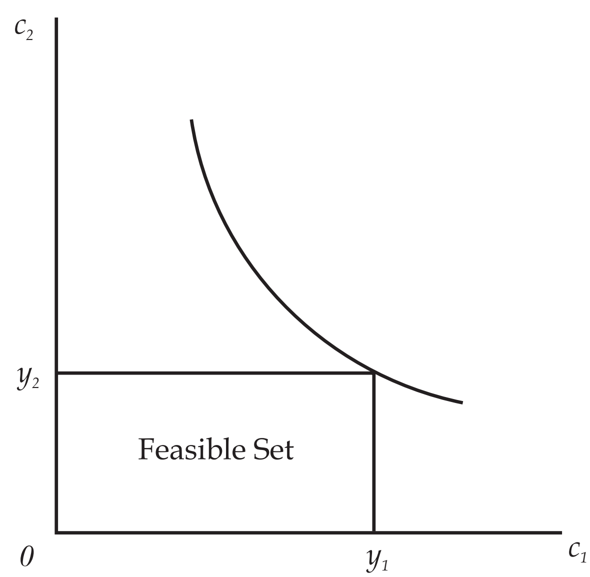
4.2.3 Robinson Crusoe

Before proceeding to develop the model further, it will be useful to pause a moment and to ask what the solution to the individual's choice problem would look like if that individual was to operate in the world just described above. In particular, assume that the individual is living alone on an island (like Robinson Crusoe, in Dafoe's famous novel). Our Robinson Crusoe is endowed with y_1 units of output (e.g., coconuts) today and expects y_2 units of output to be available in the future (coconuts that are due to mature on his trees in the next period). Remember that we are assuming that coconuts are not storable. Mathematically, the choice problem is given by:

Choose (c_1, c_2) to maximize $u(c_1, c_2)$ subject to: $c_1 \leq y_1$ and $c_2 \leq y_2$.

The solution to this choice problem is trivial: Choose $c_1^D = y_1$ and $c_2^D = y_2$. In other words, the best that Crusoe can do is to simply consume his entire income in each period. This solution is depicted diagrammatically in Figure 4.2.

FIGURE 4.2
Robinson Crusoe



4.2.4 Introducing a Financial Market

Our Robinson Crusoe economy is a *metaphor* for an environment in which individuals do not have access to markets (in this case, financial markets). In reality, individuals do have access to financial markets. A financial market is a market on which *claims* to time-dated consumption can be traded. These claims represent *promises* to deliver output at specified dates. In what follows, we assume that promises of this form can be enforced at zero cost. We implicitly made the same assumption in Chapter 2, where we thought of individuals supplying labor in exchange for claims against output; these claims were then redeemed at the firm that issued them at zero cost (i.e., there was no issue of a firm reneging on its promises). In the present context, the fact that financial claims can be costlessly enforced implies that there is no risk of default.

There are two goods in this model economy (current and future consumption) and hence there can only be one market and one price. On this market, people exchange claims for time-dated consumption and the rate at which these claims exchange measures the (relative) price of these two goods. Let R denote the price of current consumption measured in units of future consumption. That is, given some market-determined price R, an individual is able to borrow or lend one unit of c_1 in exchange for R units of c_2. For this reason, R is called the (gross) *real* rate of interest.[1] Note that the act of borrowing or lending current consumption in a private debt market is equivalent to selling or buying claims to future consumption.

- **Exercise 4.1.** You ask your buddy to buy you a beer tonight; in exchange, you promise to buy him a beer tomorrow night. What is the implicit real rate of interest in this exchange? Explain how borrowing the one beer tonight is equivalent to selling a claim against future beer.

In what follows, we shall simply assume that there is a market for risk-free private debt with an exogenously determined real interest rate R. Later on, we will discuss the economic forces that determine R, but for now we just take R as part of the environment.

If individuals have access to a financial market, then they will be faced with an *intertemporal budget constraint* (IBC). To derive this constraint, we begin by defining the concept of saving. Saving can be defined in general terms as current income minus expenditures on current needs. In the present context, saving is given by:

$$s \equiv y_1 - c_1. \tag{4.1}$$

Notice that, generally speaking, saving can be either positive or negative. If $s > 0$, then the individual is saving (purchasing bonds); if $s < 0$, then the

[1] The real interest rate is not to be confused with the nominal interest rate. The former represents the relative price of time-dated *output*, while the latter represents the relative price of time-dated *money* (a topic that will be discussed shortly).

individual is borrowing (selling bonds). Note that a 'bond' in the present context refers to a privately-issued liability (i.e., a personal promise to deliver stuff in the future).

- **Exercise 4.2.** When you approach the bank for a loan, you are in effect offering to sell the bank a bond. True, False, Uncertain and Explain.

If individuals can be expected to make good on their promises, then the second-period budget constraint is given by:

$$c_2 = y_2 + Rs. \tag{4.2}$$

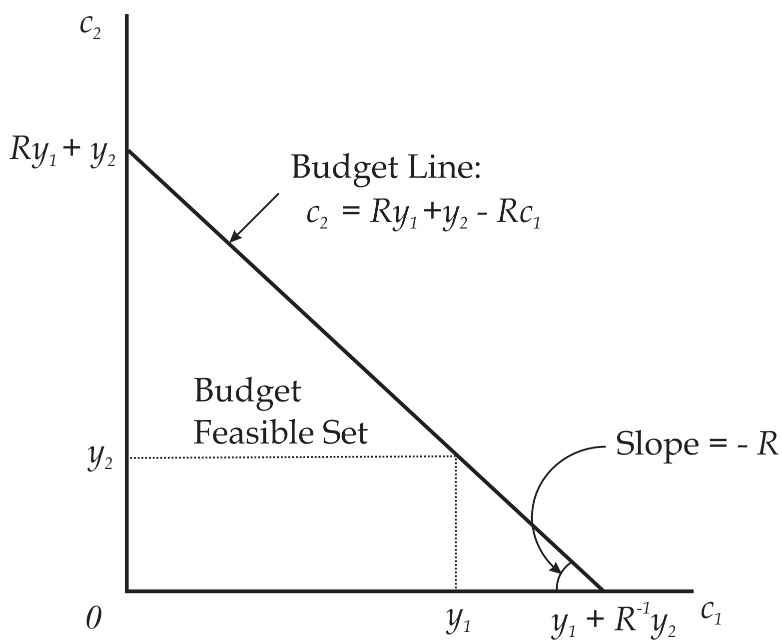
What this constraint tells us is that the individual's future consumption spending cannot exceed the sum of his earnings *plus* the interest and principal on any saving. Note that if $s < 0$, then Rs represents the amount of output that the individual is obliged to repay his creditors. Also note that since R is the *gross* interest rate, the *net* interest rate is given by $r = (R - 1)$; in other words, $(1 + r) = R$. The quantity rs is called 'interest income' if $s > 0$ and is called the 'interest charges' if $s < 0$.

Substituting the definition of saving (4.1) into the second-period budget constraint (4.2) yields an equation for the *budget line*:

$$c_2 = Ry_1 + y_2 - Rc_1. \tag{4.3}$$

This budget line tells us which combinations of (c_1, c_2) are *budget feasible*, given an endowment (y_1, y_2) and a prevailing interest rate R. Notice that the budget line is a linear function, with a slope equal to $-R$. This budget line is graphed in Figure 4.3.

FIGURE4.3
Intertemporal Budget Constraint



One can rearrange the budget line (4.3) in the following useful way:

$$c_1 + \frac{c_2}{R} = y_1 + \frac{y_2}{R}. \tag{4.4}$$

The right-hand-side of the equation above is the *present value* of the individual's lifetime earnings stream. This is just a measure of *wealth* measured in units of current consumption (and is represented as the x-intercept in Figure 4.3). We can also measure wealth in units of future consumption; i.e., $Ry_1 + y_2$. This is called the *future value* of an individual's lifetime earnings stream (and is depicted by the y-intercept in Figure 4.3). Likewise, the left-hand-side of equation (4.4) measures the present value of the individual's lifetime consumption spending. Hence, the intertemporal budget constraint tells us that the present value of lifetime consumption spending cannot exceed one's wealth. This constraint puts an upper bound on the amount that an individual can borrow.

- **Exercise 4.3.** Consider an individual with an endowment of beer given by $(y_1, y_2) = (0, 12)$. That is, the individual has no beer today, but is expecting a shipment of beer tomorrow. If the (overnight) real rate of interest is $R = 1.20$ (a 20% net interest rate), what is the maximum amount of beer that this person can borrow today?

From the intertemporal budget constraint (4.4), we see that $c_1 = y_1$ and $c_2 = y_2$ is budget feasible. Therefore, we can conclude that the IBC *always* passes through the endowment point. But unlike the case of Robinson Crusoe, we see that an individual with access to a financial market is much less constrained in how he can allocate his consumption over time. In particular, $c_1 < y_1$ and $c_2 > y_2$ is possible (by saving), or $c_1 > y_1$ and $c_2 < y_2$ is possible (by borrowing).

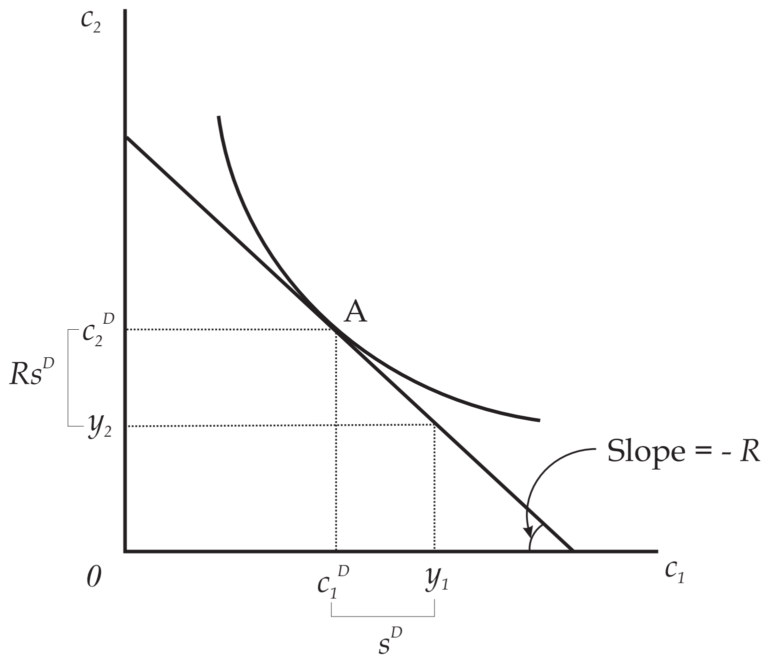
4.2.5 Individual Choice with Access to a Financial Market

For an individual in possession of an endowment (y_1, y_2) and who is able to borrow or lend at the interest rate R, the choice problem can be stated mathematically as follows:

> Choose (c_1, c_2) to maximize $u(c_1, c_2)$ subject to: $c_1 + \frac{c_2}{R} = y_1 + \frac{y_2}{R}$.

The solution to this choice problem is a pair of demand functions (c_1^D, c_2^D) that depend on the parameters that describe the person's physical and economic environment; i.e., y_1, y_2, R and u. Once c_1^D is known, one can calculate the person's desired saving function s^D from the definition of saving (4.1); i.e., $s^D = y_1 - c_1^D$. The solution is depicted graphically in Figure 4.4 as point A.

FIGURE 4.4
Consumption - Saving Choice



There are two mathematical conditions that completely describe point A in Figure 4.4. First, observe that at point A, the slope of the indifference curve is equal to the slope of the budget line. Second, observe that point A lies on the budget line. In other words,

$$MRS(c_1^D, c_2^D) = R; \tag{4.5}$$
$$c_1^D + \frac{c_2^D}{R} = y_1 + \frac{y_2}{R}.$$

- **Exercise 4.4.** Identify the exogenous and endogenous variables of the theory developed above. What sort of questions can this theory help us answer?

- **Exercise 4.5.** Suppose that the utility function takes the following form: $u(c_1, c_2) = \ln(c_1) + \beta \ln(c_2)$, where $\beta \geq 0$ is a preference parameter. Explain how the parameter β can be interpreted as a 'patience' parameter. In particular, what would β be equal to for an individual who 'doesn't care' about the future?

- **Exercise 4.6.** Suppose that preferences are such that $MRS = c_2/(\beta c_1)$. Use (4.5) to derive the consumer demand function c_1^D. How does c_1^D depend on β? Explain.

4.2.6 Small Open Economy Interpretation

Before we move on to investigate the properties of our theory, let me take some time to explain how it can be re-interpreted in a way that can help us understand some issues that arise in international trade and finance. The way we can do this is by assuming that all individuals of a domestic economy are identical.[2] Consequently, the variables c and y can be interpreted as *per capita* quantities.

An economy is said to be *open* if the residents of a domestic economy are able to trade with people inhabiting other economies. An economy is said to be *small* if the choices made by domestic residents have no measurable impact on world prices. In the context of the model developed here, the relevant 'price' is the real interest rate R. Thus, in the case of a small open economy, the interest rate prevailing in world (and hence domestic) financial markets is viewed as *exogenous*.

In international economics, the *trade balance* is defined as domestic output minus domestic spending, which in the present context is given by:

$$TB_j \equiv y_j - c_j,$$

for $j = 1, 2$. For the small open economy depicted in Figure 4.4, the economy is running a positive trade balance in the current period. That is, since domestic residents are consuming less than what they produce, the difference is exported to foreigners (in exchange for foreign bonds).[3] In the future period, the economy is running a negative trade balance (the country is a net importer of goods and services, which the country finances by 'cashing in' its accumulated foreign bond holdings).

- **Exercise 4.7.** Demonstrate that $TB_1 + \frac{TB_2}{R} = 0$. Hint: use the intertemporal budget constraint together with the definition of the trade balance.

Another important concept in international finance is called the *current account position*. The current account is defined as total income minus current expenditures. The current account differs from the trade balance only to the extent that the GNP differs from the GDP. Recall that the GNP is equal to the value of domestic production (GDP) plus *net income received from foreigners*. In a closed economy, these two values must equal, but they need not be equal if the economy is open to international trade. In the context of our model, the current account positions in periods one and two are given by:

$$CA_1 = y_1 - c_1^D;$$
$$CA_2 = y_2 + rs^D - c_2^D,$$

[2] Instructors: we could also assume that individuals have different endowments but that preferences are homothetic.

[3] The trade balance is also referred to as the *value of net exports*, denoted by NX_j in the income-expenditure identity studied in Chapter 2.

where $r = (R - 1)$.That is, since domestic residents have no foreign bonds maturing in period one (by assumption), the current account is equal to the trade balance (GDP is equal to GNP). However, in period two, domestic residents earn interest income equal to rs on the foreign bonds that are maturing in that period.[4]

Thus, in Figure 4.4, the economy is running a current account surplus. In other words, the economy is a net exporter of goods and services. Alternatively, the country is a net importer of foreign bonds, so the country is also running a *capital account deficit* (the capital account is simply defined as the negative of the current account). Notice that $CA_1 = s^D$. For this reason, the current account in this model economy corresponds to *net domestic saving*. Note that in the future period, $CA_2 = -s^D$ (a country that saves in the current period will dissave in the future period, and vice-versa).

4.3 Experiments

The model developed above constitutes a theory of consumer demand (and saving). Alternatively, in the context of a small open economy, the theory explains the determination of aggregate consumer spending and the current account. This theory takes the following form:

$$(c_1^D, c_2^D, s^D) = f(y_1, y_2, R, u).$$

Note that one of the benefits of being explicit about the intertemporal aspects of decision-making is that we can make a precise distinction between the effects of *transitory*, *anticipated*, and *permanent* changes in the physical and economic environment. For example, note that our theory asserts that current consumer demand should depend not only on current income, but also on the level of income that is expected in the future.

4.3.1 A Transitory Increase in Current GDP

In Chapter 2, we learned how a positive productivity shock could lead to an increase in GDP. Imagine here that this productivity shock is transitory (temporary) so that $\Delta y_1 > 0$ and $\Delta y_2 = 0$. Because this is a small economy, this productivity shock has no effect on R. How do the people living in our model economy react to such a development?

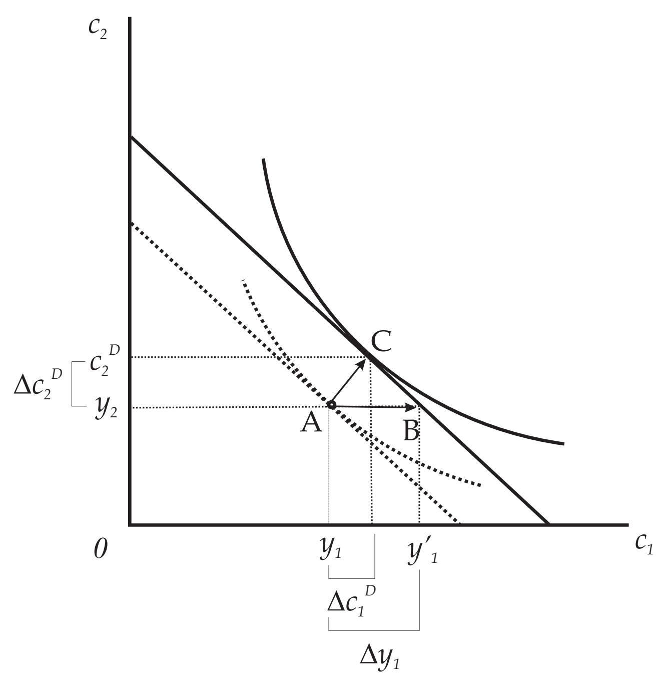
We can answer this question with the aid of a diagram. The first step is to depict the pattern of desired consumption and saving just prior to the shock; this situation is depicted as point A in Figure 4.5. I have drawn point A such that

[4] In the present context, you can think of $s^D > 0$ as the value of domestic capital that is 'working' in some foreign country earning an income rs^D. Likewise, you can think of $s^D < 0$ as the value of foreign capital employed on domestic soil earning foreigners an income of rs^D.

the country is initially running a zero trade balance, but nothing important that I say below depends on this (feel free to begin with either a positive or negative trade balance). Now, suppose that $\Delta y_1 > 0$. Since $\Delta y_2 = 0$, we can depict this shift as a rightward shift of the endowment (A $\rightarrow$ B). Since the interest rate is unaffected, this implies a rightward shift of the intertemporal budget constraint. Note that the shock has made domestic residents *wealthier*.

The question now is where to place the new indifference curve. If we make the reasonable assumption that consumption at each date is a *normal good*, then the increase in wealth results in an increase in consumer demand in *both* periods; i.e., $\Delta c_1^D > 0$ and $\Delta c_2^D > 0$. We can depict such a response by placing the new indifference curve at a point northeast of the original position; e.g., point C in Figure 4.5.

FIGURE 4.5
A Transitory Increase in GDP



So we see that a transitory productivity shock results in a relatively mild but prolonged 'consumption boom.' Notice that the increase in current consumer demand is less than the increase in current GDP; i.e., $\Delta c_1^D < \Delta y_1$. The ratio $\Delta c_1^D / \Delta y_1$ is called the *marginal propensity to consume out of current income*. Since $\Delta c_1^D / \Delta y_1 < 1$, we see that a one dollar increase in income results in a *less than* one dollar increase in current consumer demand when the income shock is transitory. The extra income that is not consumed is saved. In this model, the extra saving takes the form of purchases of foreign bonds (hence, the country moves to a current account surplus). This increase in saving is used to finance the higher consumption level that is desired in the future.

The assumption that consumption at each date is a normal good can be interpreted as a preference for *consumption-smoothing*. That is, any increase in wealth will be spread over all periods in the form of higher consumption at every date. The availability of a financial market allows individuals to smooth their consumption over time in response to transitory changes in their income (contrast this with how Robinson Crusoe would have to react to a similar shock). As such, one can think of financial markets as supplying a type of 'shock-absorber' against transitory income shocks. That is, by saving (or borrowing), individuals can use financial markets to absorb the impact of transitory income shocks and in this way keep their lifetime consumption patterns relatively stable.

- **Exercise 4.8.** Calculate the marginal propensity to consume for Robinson Crusoe.

- **Exercise 4.9.** Imagine modifying our model to allow for many time periods. Explain why, for any given income shock (e.g., $\Delta y_1 = \$1000$), the response of current consumer spending is likely to grow smaller as the time horizon expands.

- **Exercise 4.10.** In a recent article, John Bluedorn (2002) investigates how the current account position of small Caribbean and Central American economies react to 'hurricane shocks.' Hurricanes are not infrequent events in these parts of the world. When they hit, they invariably lead to a transitory decline in real per capita GDP. The author finds that the current account position of these economies first falls and later increases in response to a hurricane shock. Is this feature of the data consistent with our theory? Explain.

- **Exercise 4.11.** A newspaper article reports the following event with commentary: "The Canadian economy has just entered a recession (a transitory decline in GDP). Adding to our problems is the ballooning current account deficit." Explain why the 'ballooning' current account deficit is likely a 'good' thing. Hint: model the recession as an exogenous $\Delta y_1 < 0$ and evaluate economic welfare under two scenarios: one in which the current account moves into deficit; and one in which the government prevents domestic residents from selling bonds to foreigners (so that the current account position remains in balance).

4.3.2 An Anticipated Increase in Future GDP

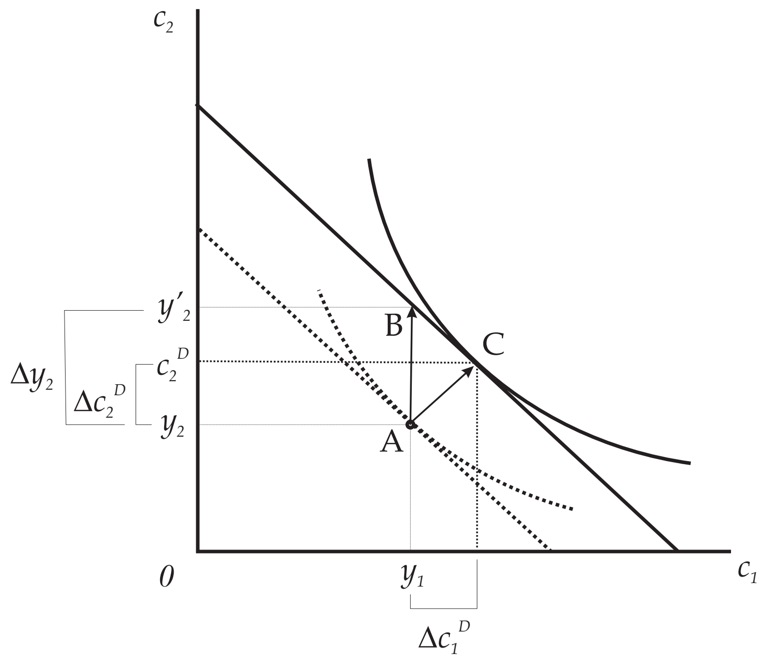
Imagine now that the residents of our model economy receive news (information) that leads them to revise upward their forecast of their future earnings. Since the sum of these earnings is equal to y_2, we can model such a shock as

$\Delta y_1 = 0$ and $\Delta y_2 > 0$. An example of such an 'information shock' could be what typically seems to happen just before an economy emerges from recession. Alternatively, one could imagine the arrival of a new technology that is expected to improve GDP in the near future. How would our model individuals react to such information?

Again, we can answer this question with the aid of a diagram. The first step is to depict the pattern of desired consumption and saving just prior to the shock; this situation is depicted as point A in Figure 4.6. Once again, I have drawn point A such that the country is initially running a zero trade balance (again, feel free to begin with either a positive or negative trade balance). Now, suppose that $\Delta y_2 > 0$. Since $\Delta y_1 = 0$, we can depict this shift as an upward shift of the endowment (A $\rightarrow$ B). Since the interest rate is unaffected, this implies an upward shift of the intertemporal budget constraint. Note that while the shock leaves current GDP unchanged, it nevertheless makes domestic residents *wealthier*. This example makes it clear why it is important to distinguish between *income* and *wealth*.

The question now is where to place the new indifference curve. Assuming that consumption at each date is a normal good, then the increase in wealth results in an increase in consumer demand in *both* periods; i.e., $\Delta c_1^D > 0$ and $\Delta c_2^D > 0$. We can depict such a response by placing the new indifference curve at a point northeast of the original position; e.g., point C in Figure 4.6.

FIGURE 4.6
An Anticipated Increase in GDP



Here, we see that the anticipated increase in future GDP also results in a 'consumption boom' that begins in the *current* period. The intuition for this is the same as before: the shock results in a higher level of wealth so that the consumption-smoothing motive implies that desired consumer spending in all periods rises. However, note that while consumption responds in a manner similar to when the economy is hit by a transitory shock, the behavior of savings is quite different. In particular, this shock causes a *decline* in domestic saving (so that the economy moves into a current account deficit position). Anticipating their higher future earnings, domestics increase their current consumption by borrowing from (selling bonds to) foreigners. Once again, observe how the availability of a financial market allows individuals to smooth their consumption in the face of an income shock.

The example portrayed in Figure 4.6 also reveals another important point. Notice that while the current account position of this economy has 'deteriorated' (to use language that is common in the financial pages of newspapers), the welfare of domestic residents is *higher* than before. Consequently, we can conclude that we must be careful in drawing any immediate link between a country's current account position and the welfare of its residents.

- **Exercise 4.12.** You receive some 'good' news that your Aunt has just passed away and left you with a huge inheritance. Unfortunately, you are able to collect on this inheritance only once you graduate (in the near future, hopefully). Since you are currently a student, your current income is rather low and you are subsisting largely on Kraft dinners. Explain what action you could take to increase your current spending. Assume that the fact of your future inheritance is perfectly verifiable (by a bank manager, for example).

An interesting feature of real economies emerging from recession is that consumer spending often recovers before GDP does. This empirical observation is often interpreted as evidence that an increase in consumer spending 'causes' economic growth. According to our theory, however, the direction of causality actually works in reverse. That is, the increase in consumer spending today is caused by the *arrival of information* that leads individuals to *revise upward their forecast of future income*. For example, laid off workers may receive information that their former employers are planning to rehire in the immediate future. To the extent that individuals are on average correct in their forecasts, the increase in consumer spending will preceed the rise in aggregate income.

The example above warns us to be careful in trying to infer causality simply by looking at correlations in the data. Correlations by themselves are nothing more than measurements (descriptions) of the data; they do not constitute theory. Any particular intertemporal correlation may in fact be generated by what econometricians call *reverse causality*. To better understand the concept of reverse causality, think about the behavior of consumers during the Christmas season. It is an empirical fact that Christmas shopping preceeds Christmas. However, it would be wrong to conclude on the basis of this correlation that Christmas shopping causes Christmas. The direction of causality is obviously reversed.

- **Exercise 4.13.** Let preferences be such that $MRS = c_2/(\beta c_1)$. Solve for the desired saving rate s^D/y_1 and explain how the behavior of the saving rate could be used by econometricians to forecast economic growth (y_2/y_1).

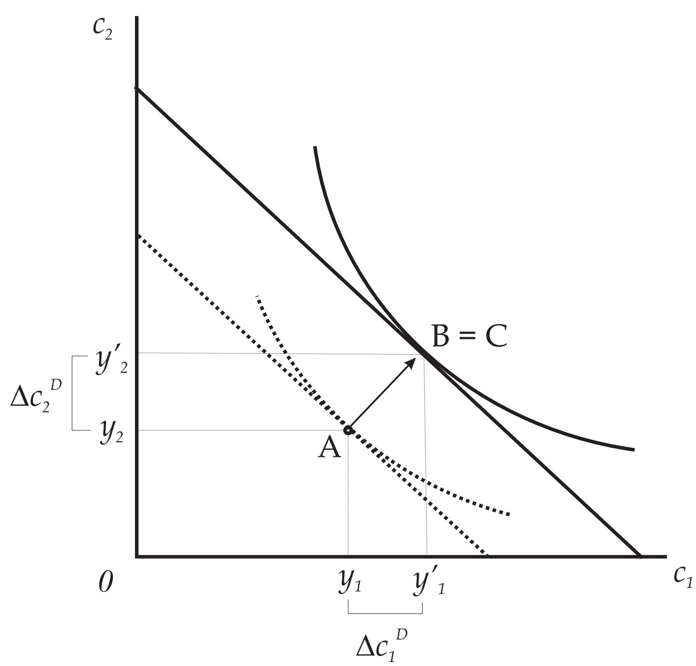
4.3.3 A Permanent Increase in GDP

Imagine now that the economy experiences a productivity shock that is expected to be permanent. A permanent productivity shock can be modeled here as $\Delta y_1 = \Delta y_2 = \Delta y > 0$. Notice that a permanent shock to GDP is a combination of the two shocks studied above.

Again, we can answer this question with the aid of a diagram. The first step is to depict the pattern of desired consumption and saving just prior to the shock; this situation is depicted as point A in Figure 4.7. Once again, I have drawn point A such that the country is initially running a zero trade balance (again, feel free to begin with either a positive or negative trade balance). Now, since $\Delta y_1 = \Delta y_2 = \Delta y > 0$, we can depict this change as a 45^0 shift of the endowment (A $\rightarrow$ B). Since the interest rate is unaffected, this implies an outward shift of the intertemporal budget constraint. Once again, the shock makes individuals wealthier. Note that the increase in wealth is *greater* than the case in which the shock to GDP was transitory.

The question now is where to place the new indifference curve. Assuming that consumption at each date is a normal good, then the increase in wealth results in an increase in consumer demand in *both* periods; i.e., $\Delta c_1^D > 0$ and $\Delta c_2^D > 0$. Notice that the shift in the consumption pattern is similar to the shift in the endowment pattern. While this shift need not be precisely identical, for simplicity assume that it is. In this case, $\Delta c_1^D = \Delta y$ and $\Delta c_2^D = \Delta y$. We can depict such a response by placing the new indifference curve at a point northeast of the original position; e.g., point C in Figure 4.7.

FIGURE 4.7
A Permanent Increase in GDP



Once again, the consumption response is similar to the other two experiments. Note, however, that the size of the increase in consumer spending is much larger here, compared to when the income shock was transitory. In particular, our theory predicts that the marginal propensity to consume out of current income, when the income shock is perceived to be permanent, is (approximately) equal to $\Delta c_1^D / \Delta y_1 = 1.0$. In other words, our theory suggests that the marginal propensity to consume out of current income *depends critically on whether shocks to income are perceived to be transitory or permanent.*

4.3.4 A Change in the Interest Rate

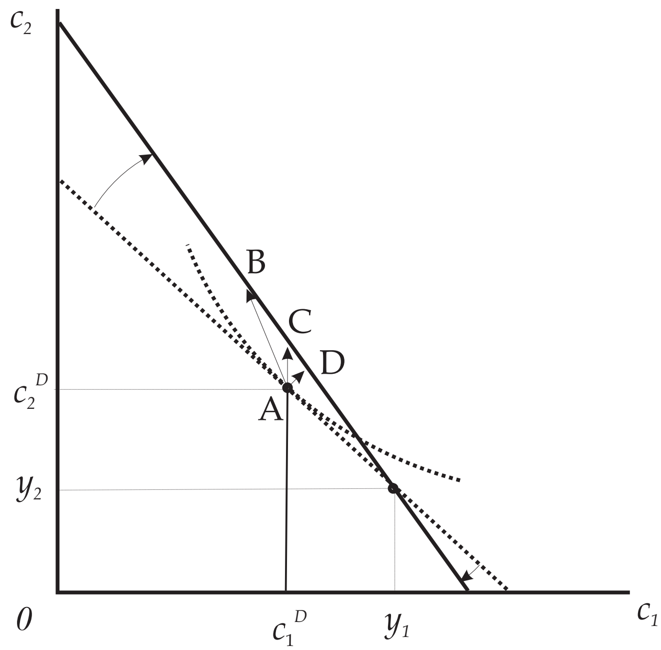
A change in the interest rate changes the slope of the intertemporal budget constraint, which implies a change in the relative price of current and future consumption. Whenever a price changes, we know that in general there will be both a substitution effect and a wealth effect at work, making the analysis slightly more complicated. As it turns out, what we can say about how individuals react to a change in the interest depends on whether the individual is planning to be a borrower or a lender. We will consider each case in turn.

Lenders

Individuals planning to lend are those people who currently have high income levels but are forecasting a decline in their future income; i.e., $y_1 > y_2$. Individuals who are in their peak earning years (and thus approaching retirement age) constitute a classic example of people who generally wish to save. Point A in Figure 4.8 depicts the case of a lender. If the interest rate rises, then current consumption becomes more expensive than future consumption. The substitution effect implies that people would want to substitute out of c_1 and into c_2. This applies to both borrowers and lenders. What will differ between the two cases is the wealth effect.

Observe that the effect of an increase in the interest rate on wealth depends on how wealth is measured. That is, wealth measured in present value declines, but wealth measured in future value rises. For a lender, it is appropriate to think of wealth as increasing with the interest rate. The intuition for this is that when R rises, the value of current output rises and lenders are those people who are relatively well endowed in current output. Consequently, the wealth effect for a lender implies that both c_1 and c_2 increase. Notice that while the substitution and wealth effects operate in the same direction for c_2, we can conclude that c_2^D unambiguously rises. However, the substitution and wealth effects on c_1 operate in opposite directions. Thus, c_1^D may either rise or fall, depending on the relative strengths of these two effects. Nevertheless, we can conclude that an increase in the interest rate leads to an unambiguous increase in welfare for lenders.

FIGURE 4.8
An Increase in the Interest Rate
(Lenders)



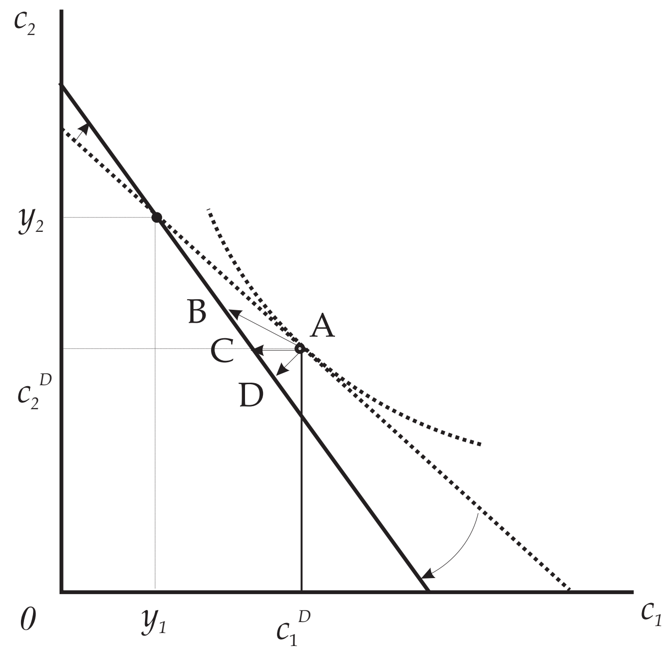
Borrowers

Individuals planning to borrow are those who currently have low income levels but are forecasting higher incomes in the future (i.e., $y_1 < y_2$). Young individuals approaching their peak earning years (e.g., university students) constitute a classic example of people who generally wish to borrow. Point A in Figure 4.9 depicts the case of a borrower.

The substitution effect associated with an increase in the interest rate works in the same way as before: Individuals would want to substitute out of the more expensive good (c_1) into the cheaper good (c_2). The difference here, relative to the case of a lender, is in the wealth effect. For a borrower, an increase in the interest rate lowers the value of the good that borrowers are relatively well endowed with (future income). Consequently, they are made less wealthy. This reduction in wealth leads to a decline in both c_1 and c_2.

Note that the substitution and wealth effect now operate in the same direction with respect to c_1. Consequently, we can conclude that an increase in the interest rate leads those who are planning to borrow to scale back on their borrowing (i.e., increase their saving), so that c_1^D unambiguously declines. On the other hand, the substitution and wealth effects operate in opposite directions with respect to c_2. Therefore, c_2^D may either rise or fall depending on the relative strength of these two effects. In any case, it is clear that borrowers are made worse off (they are on a lower indifference curve) if the interest rate rises.

FIGURE 4.9
An Increase in the Interest Rate
(Borrowers)



Of course, everything said here can also apply to a small open economy. In particular, how a small open economy responds to change in the world interest rate depends on whether the country is a net creditor or a net debtor nation.

4.4 Borrowing Constraints

The analysis in this chapter assumes that individuals are free to borrow or lend at the market interest rate. However, in reality, this may not always be the case. In particular, it is not clear that those wishing to borrow (with the willingness and ability to pay back their debt) can always do so. Likewise, a country that wishes to borrow may not always be able to obtain the credit that is desired. The reasons for why this may be the case are varied, but to the extent that it is true, then borrowers are said to face *borrowing constraints* that limit the amount that can be borrowed.

A skeptic may remark that the world is full of people (and countries) that would like to 'borrow,' while having little intention of paying back their debt. Or perhaps the intention is there, but some individuals may be overly optimistic concerning their ability to repay. The point here is that, in practice, it is difficult to know whether some individuals are truly debt-constrained or whether they would in fact be violating their intertemporal budget constraint. The challenge for theorists is to explain why creditors would refuse to lend to people (or countries) who are in a position to make good on their promise to repay.

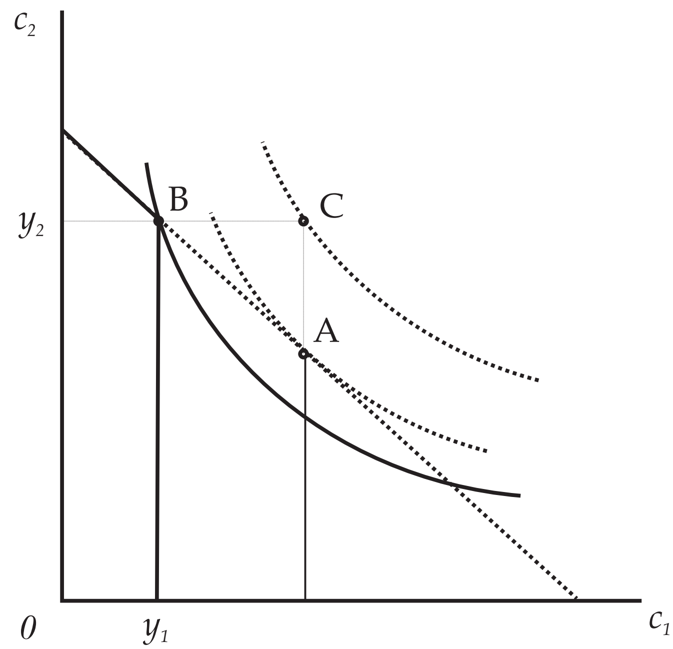
One way to think about borrowing constraints is as follows. Every loan requires *collateral* in one form or another. Collateral is an asset that serves to back a loan and measures the ability (not necessarily the willingness) of an individual to back up promises to repay. In the context of our model, the collateral for a loan is given by an individual's (or country's) future income y_2. If an individual could pledge y_2 as collateral, then the individual would have no problem in borrowing up to the present value of his collateral; i.e., y_2/R.

But if y_2 represents future labor earnings, then there may be a problem in securing debt by pledging y_2 as collateral. In particular, most governments have passed laws that prevent individuals from using future labor income as collateral. These restrictions are reflected in laws that make human capital *inalienable*.[5] What this means is that if an individual borrows resources from a creditor, then the creditor is legally prohibited from seizing that individual's future labor income in the event that the individual refuses to repay his debt. In effect, the debtor is legally prohibited from using future labor income as collateral. For example, personal bankruptcy laws allow individuals to discharge their debt (to private creditors, not government creditors) with virtual impunity. Understanding this, a rational creditor is unlikely to extend a loan, even though the debtor has the ability to repay. The same holds true for countries. The only way to force a nation in default of its loans to repay would be through an act of war. Understanding this, international creditors may be unwilling to extend loans to countries with a poor record of repayment, even if the debtor nation technically has the means to repay its loans.

We can use a familiar diagram to display the effects of borrowing constraint. Every individual continues to face an intertemporal budget constraint $c_1 + R^{-1}c_2 = y_1 + R^{-1}y_2$. Suppose, however, that individuals are free to save but not borrow. In this case, individuals face an additional constraint: $c_1 \leq y_1$ (they cannot consume more than they earn). Point A in Figure 4.10 displays the case of a borrower who is able to borrow. Point B shows where this individual must consume if he is subject to a borrowing constraint.

[5] See Andolfatto (2002).

FIGURE 4.10
Borrowing Constraints



If the borrowing constraint is binding (i.e., if the individual is at point B), then two things are immediately clear. First, the individual is clearly worse off relative to the case in which he is able to borrow (point A). Second, the marginal propensity to consume out of current income for individuals who are debt constrained is equal to one (even for transitory income shocks).

Now, let us consider the following interesting experiment. Consider an economy populated by a current generation of students (with endowment given by point B in Figure 4.10). Suppose that initially, these students are free to borrow at interest rate R, so that they attain the point A in Figure 4.10. Clearly, these students are racking up a lot of student debt. Suppose now that these students (or their representatives) lobby the government, complaining about their 'unfair' levels of debt and how unreasonable it is to expect them to repay it. Bowing to this pressure, the government passes a law that allows students to default on their debt. Judging by the high incidence of student debt default in reality, many students appear willing to take up such an option. By defaulting on their debt, these students move from point A to point C in Figure 4.10. Clearly, these students are made better off (at the expense of their creditors – those evil banks that are owned by...their parents?).

But while the current generation of students is made better off by such a law, the same cannot be said of future generations of students. In particular, creditors who are burned by the law are unlikely to make the same mistake twice. Creditors would refuse to extend new loans to new generations of students. These generations of students must consume at point B, instead of point A.

The preceeding discussion raises many interesting questions. In particular, it seems clear enough that even though individual labor income cannot be used as collateral, many individuals are apparently both willing and able to obtain large amounts of 'unsecured' consumer debt. Of course, some of this debt is subject to default. However, most of it is repaid. The question is *why*? Similarly, while some nations (and local governments) occasionally default on their debt obligations, most do not. Again, the question is *why*? An obvious reason may be that by developing a good credit history, an individual (or country) can ensure that he (it) has access to credit markets in the future. Appendix 4.A provides a real world example of this principle at work.

4.5 Determination of the Real Interest Rate

Thus far, we have simply assumed that the real rate of interest was determined exogenously (e.g., given by God or Nature). As far as individuals (or small open economies) go, this seems like an appropriate assumption to make, since if decision-making agents are small relative to the world economy, then their individuals actions are unlikely to affect market prices. That is, from an individual's perspective, it is 'as if' market prices bounce around exogenously according to some law of nature.

But it remains true that the real interest rate is a *market* price and that market prices are determined, in part, by the behavior of individuals *collectively*. In other words, while it may make sense to view some things as being exogenous to the world economy (e.g., the current state of technical knowledge), it does not make sense to think of a market price in the same way. It makes more sense to think of market prices as being determined *endogenously* by *aggregate* supply and demand conditions.

In order to think about what determines the real rate of interest, we will have to think of things in terms of the *world* economy, or at the very least, a large open economy (like the United States). Unlike a small open economy (e.g., individuals or small countries), the world economy is a *closed* economy. Thus, while it may make sense for an individual country to run a current account surplus (or deficit), it does not make sense for all countries to run a surplus (or deficit) simultaneously (unless you believe that some world citizens are trading with aliens). As far as the world is concerned, the current accounts of all countries together *must sum to zero.*

- **Exercise 4.14.** You and your friend Bob are the only two people on the planet. If you borrow a case of beer (at zero interest) from Bob and promise to pay him back tomorrow, then describe the intertemporal pattern of individual and aggregate current account positions in this economy.

A closed economy model is sometimes referred to as a *general equilibrium* model. A general equilibrium model is a model that is designed to explain the determinants of market prices (as well as the pattern of trade). In contrast, a small open economy is a model in which market prices are viewed as being exogenous. Such models are sometimes referred to as *partial equilibrium* models, since while they are able to explain trade patterns as a function of the prevailing price-system, they do not offer any explanation of where these prices come from.

4.5.1 General Equilibrium in a 2-Period Endowment Economy

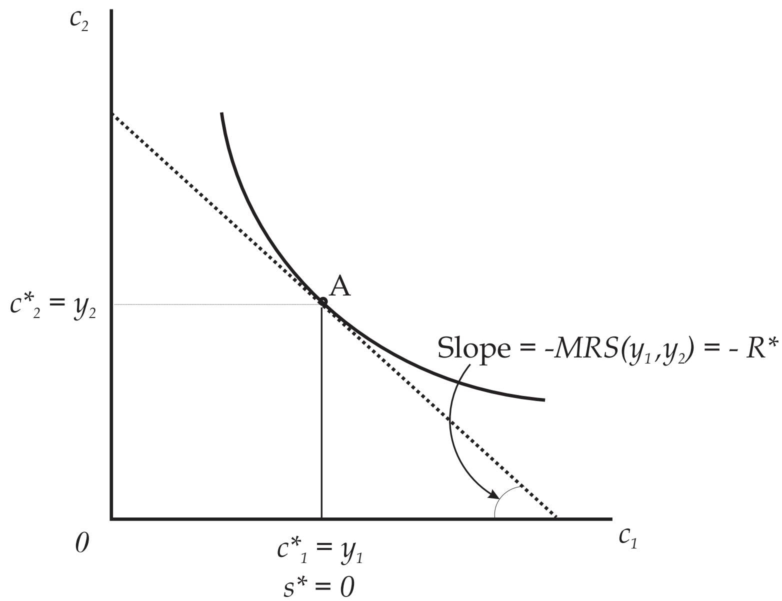
Consider Figure 4.4. This figure depicts an individual's desired consumption (and saving) profile *given* some intertemporal pattern of earnings (y_1, y_2) and *given* some (arbitrary) real rate of interest R. In this section, we will continue to view (y_1, y_2) as exogenous (which is why we call this an *endowment* economy). But we now ask the question: "How is R determined...where does it come from?"

In order to answer this question, we will have to reinterpret Figure 4.4 as depicting the world economy. That is, let us now interpret (c_1, c_2) as the consumption profile of a 'representative agent' and (y_1, y_2) as the intertemporal pattern of real per capita output in the *world* economy. Figure 4.4 then continues to depict a partial equilibrium. That is, given some arbitrary real rate of interest R, the 'average' world citizen desires to save some positive amount; i.e., $s^D > 0$.

But clearly, $s^D > 0$ cannot be a general equilibrium. That is, it is impossible for the world's net credit position to be anything other than zero. The partial equilibrium depicted in Figure 4.4 features an excess supply of loanable funds (excess desired savings). This is equivalent to saying that there is an excess supply of current output $(c_1^D < y_1)$ or an excess demand for future output $(c_2^D > y_2)$. In this model, everyone wants to save and nobody wants to borrow given the prevailing rate of interest. Something has to give. It seems natural, in the present context, to suppose that what has to 'give' here is the prevailing rate of interest. In particular, the excess supply of loanable funds is likely to drive the market interest rate down (the converse would be true if there was an excess demand for credit).

Since the net value of consumption loans must be equal to zero, it seems natural to suppose that the real rate of interest will adjust to the point at which $s^D = 0$. Note that when $s^D = 0$, we also have $c_1^D = y_1$ and $c_2^D = y_2$. Let R^* denote the equilibrium real rate of interest; that is, the rate of interest that sets $s^D = 0$. This equilibrium interest rate is depicted in Figure 4.11.

FIGURE 4.11
General Equilibrium



Notice that in Figure 4.11, individuals are still thought of as viewing the prevailing interest rate R^* as exogenous with respect to their own personal decisions concerning how much to consume and save. In (general) equilibrium, however, the interest rate must adjust so that all individual decisions are *consistent* with each other. Since everyone is the same in this simple model, logic dictates that the only consistent savings decision is for everyone to choose $s^D = 0$. The only interest rate that will make $s^D = 0$ an optimal choice is R^*.[6]

In this simple endowment economy, total (world) consumption must be equal to total (world) output; i.e., $c_1 = y_1$ and $c_2 = y_2$. Since individuals are optimizing, it must still be the case that $MRS = R^*$ (notice that the slope of the indifference curve in Figure 4.11 is tangent to the intertemporal budget constraint exactly at the endowment point). Suppose that preferences are such that $MRS = c_2/(\beta c_1)$, where $0 < \beta < 1$. Then since $c_1 = y_1$ and $c_2 = y_2$ (in equilibrium), our theory suggests that the equilibrium real rate of interest is given by:

$$R^* = \frac{1}{\beta}\left(\frac{y_2}{y_1}\right). \tag{4.6}$$

[6] The analysis here easily extends to the case of many different individuals or economies. That is, consider a world with N different countries. Then, given R^*, it is possible for $s_i^D \gtrless 0$ for $i = 1, 2, ..., N$ as long as $\sum_{i=1}^{N} s_i^D = 0$.

Equation (4.6) tells us that, in theory, the real rate of interest is determined in part by preferences (the patience parameter β) and in part by the expected growth rate of the world economy (y_2/y_1). In particular, theory suggests that an increase in patience (β) will lead to a lower real rate of interest, while an increase in the expected rate of growth (y_2/y_1) will lead to a higher real rate of interest. Let us take some time now to understand the intuition behind these results.

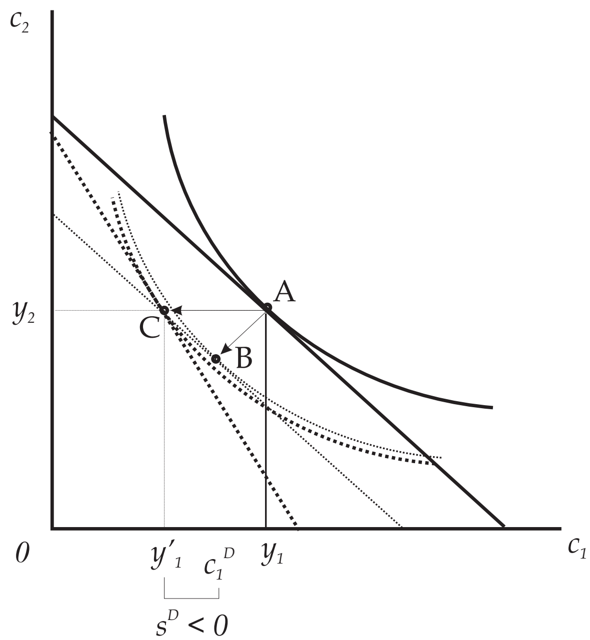
4.5.2 A Transitory Decline in World GDP

Imagine that world output falls unexpectedly below its trend level so that $\Delta y_1 < 0$ (the world economy enters into a recession). Imagine furthermore that this recession is not expected to last very long, so that $\Delta y_2 = 0$. Since the recession is expected to be transitory (short-lived), the unexpected drop in current world GDP must lead to an *increase* in the expected rate of growth (y_2/y_1) as individuals are forecasting a quick recovery to 'normal' levels of economic activity. What sort of effect is such a shock likely to have on the real rate of interest?

According to our theory, any shock that leads individuals to revise their growth forecasts upward is likely to put upward pressure on the real rate of interest. The intuition behind this result is straightforward. Since real incomes are perceived to be low for only a short period of time, standard consumption-smoothing arguments suggest that individuals will want to reduce their desired saving (increase their desired borrowing), thereby shifting a part of their current burden to the future. If the interest rate was to remain unchanged, then in aggregate there would be an excess demand for credit (too few savers and too many borrowers); i.e., $s^D < 0$. In a competitive financial market, one would expect the excess demand for credit to put upward pressure on the interest rate. In equilibrium, the interest rate must rise to the point where once again $s^D = 0$.

Figure 4.12 depicts this experiment diagrammatically. Imagine that the initial equilibrium is at point A. A surprise decline in current world output moves the world endowment to point C. If we suppose, for the moment, that the interest rate remains unchanged, then consumption-smoothing behavior moves the desired consumption profile to point B. At point B, however, there is an excess demand for current period consumption; i.e., $c_1^D > y_1'$, or equivalently, an excess demand for credit; i.e., $s^D < 0$. In order to eliminate the excess demand for credit, the real interest rate must rise so that the credit market clears; this occurs at point C.

FIGURE 4.12
A Transitory Recession Leads to an
Increase in the Real Rate of Interest



- **Exercise 4.15.** Using a diagram similar to Figure 4.12, show that an increase in the expected growth rate of world GDP brought about by news that leaves current GDP unchanged, but leads to an upward revision for the forecast of future GDP, also leads to an increase in the real rate of interest. Explain.

4.5.3 A Persistent Decline in World GDP

As with individual economies, the growth rate in world real GDP fluctuates over time. Any given change in the growth rate may be perceived by market participants as being either *transitory* (e.g., lasting for a year or less) or *persistent* (e.g., possibly lasting for several years). In the previous subsection, we considered the case of a transitory increase in the expected rate of growth (brought about by a transitory decline in the current level of world GDP). There may be other circumstances, however, in which a change in the rate of growth is perceived to be longer lasting (persistent). Extended periods of time in which growth is relatively low (not necessarily negative) are called *growth recessions*.

Let us now consider the following experiment. Imagine that current GDP is unexpectedly low, as in the previous experiment. But unlike the previous experiment, let us now imagine that individuals perceive that the growth rate is expected to fall. In particular, imagine that the $\Delta y_1 < 0$ leads individuals to revise downward their growth forecasts so that $\Delta(y_2/y_1) < 0$. According to our theory, such an event would lead to a decline in the real rate of interest. The intuition for this is relatively straightforward. That is, even though current GDP declines, future GDP is forecast to decline by even more, leading individuals to increase their desired saving (reduce their desired borrowing). The excess supply of loanable funds puts downward pressure on the real rate of interest. The opposite would happen if financial markets suddenly received information that led participants to revise upward their forecasts of world economic growth.

To summarize, our theory suggests that a short-term rise in the real interest rate is likely to occur in the event of a (perceived) short-term decline in the *level* of GDP below trend. On the other hand, to the extent that a recession takes the form of lower expected growth rates (expected persistent declines in the level of GDP below trend), the real rate of interest is likely to fall. Conversely, a world economic boom that takes the form of higher expected growth rates is likely to result in higher real rates of interest.

4.5.4 Evidence

Figure 4.13 plots the *actual* growth rate in real GDP for the United States against a measure of the short-term (one year) real interest rate.[7] Since the United States is a large economy, it seems reasonable to suppose that movements in this (large) economy are highly correlated with movements in world variables. According to our theory, the short-term real interest rate should fluctuate in accordance with the market's *expectation* of short-term real growth in GDP. Unfortunately, measuring the market's expectation of future growth is not a straightforward task, making it difficult to test our theory. In the absence of data on market expectations, the theory can nevertheless be used as an *interpretive* device.

From Figure 4.13, we see then that the real interest rate is not a very good predictor of future growth. Perhaps this is because forecasting future growth rates is an inherently difficult exercise for market participants. Note that the real rate of interest was very low (even negative) in the mid-1970s. According to our theory, market participants were expecting the economic contraction in 1974-75 to last longer than it did. Likewise, note the unusually high interest rates that occurred during the contractions in the early 1980s. Our theory suggests that market participants were surprised by the length of the slowdown in economic growth. On the other hand, both real interest rates and growth rates were high during the late 1980s and the late 1990s. In these cases, it appears that market participants correctly anticipated these periods of economic boom. Finally, note

[7] The real interest rate measure here was computed by taking the nominal yield on one-year U.S. government securities and substracting the one-year ahead forecast of inflation based on the Livingston survey; see: www.phil.frb.org/ econ/ liv/

that according to more recent data, the real interest rate is again in negative territory, while economic growth appears to be relatively robust. Evidently, the market is still expecting some short-term weakness in the U.S. economy. Whether these expectations are confirmed remains to be seen.

Figure 4.13
Growth Rate in Real per Capita GDP (Actual)
and the Real Short-Term Rate of Interest
United States (1970.1 - 2000.3)

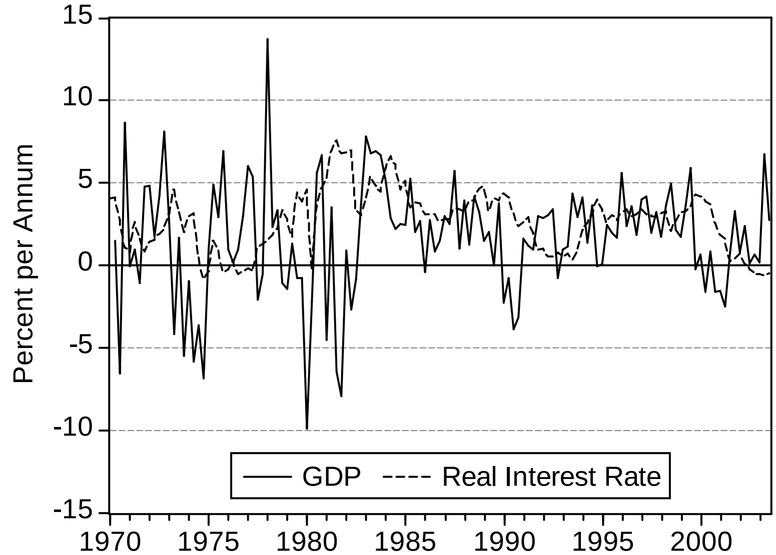


Figure 4.14 plots an estimate of the growth rate in (total) world real GDP.[8] As argued above, the (expected) growth rate in world GDP is likely a better measure to use (especially as capital markets become increasingly integrated). Unfortunately, there is no readily available measure of the real world interest rate. However, Figure 4.15 plots a measure of the short-term (*ex post*) real interest rate, which is based on the U.S., Euro area, and Japanese economies.

[8] These numbers are based on Madison's estimates; see: www. theworldeconomy.org/statistics.htm

Figure 4.14
World Real GDP Growth
1970 - 2001

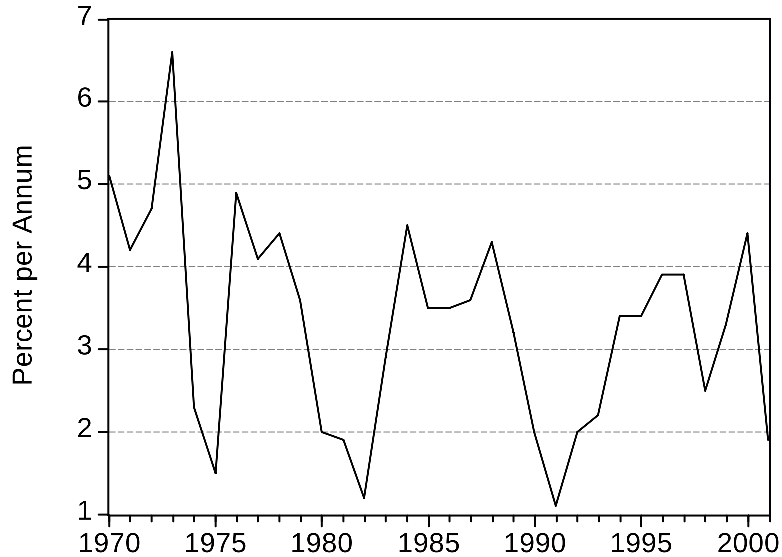
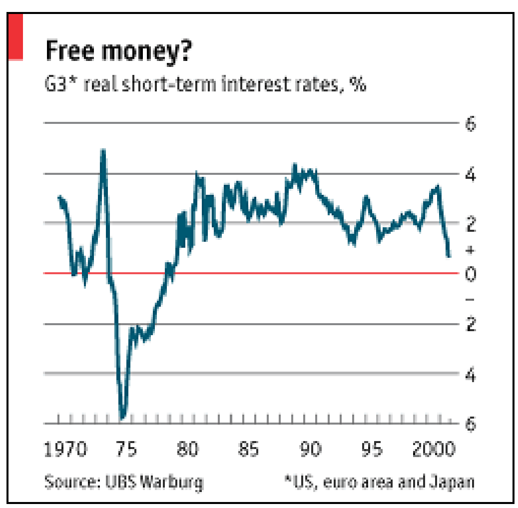


Figure 4.15



The striking feature in Figure 4.15 are the very low rates of return realized in the mid-1970s. Indeed, world growth did turn out to be lower than average during this period of time. Since the early 1980s, the real interest rate has fluctuated between one and four percent, tending to fall during periods of slow growth and tending to rise (or remain stable) during periods of more rapid growth.

4.6 Summary

Many, if not most, decisions involve an intertemporal dimension. Actions today can have implications for the future. Any act of saving is necessarily dynamic in nature. By saving more today, an individual (or country) can consume more tomorrow. Since saving more today implies less consumption today (for a given stream of income), the saving decision is related to the choice of how to allocate consumption expenditures over time. In other words, consumer demand should also be thought of as involving a dynamic dimension.

With the availability of financial markets, individuals (or small open economies) are no longer constrained to live within their means on a period-by-period basis. Instead, they are constrained to live within their means on a lifetime basis. As such, financial markets provide a type of 'shock absorber' for individuals; allowing them to smooth their consumption in the face of shocks to their income. As a corollary, it follows that desired consumer spending at any point in time is better thought of as depending on the wealth of the household sector, rather than on income. Shocks to income may influence consumer spending, but only to the extent that such shocks affect wealth. From this perspective, it also follows that the impact of income shocks on consumer demand can depend on whether such shocks are perceived to be transitory or persistent.

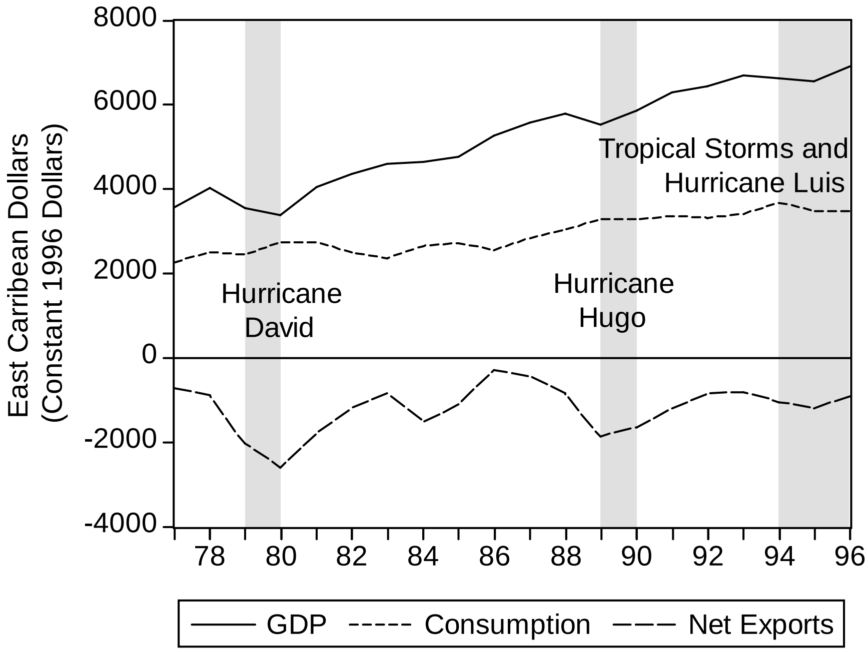
From the perspective of an open economy, aggregate saving is related to a country's current account position (or trade balance). A current account surplus is simply a situation where total domestic income exceed total domestic consumer spending. This difference must therefore reflect the value of net exports. The converse holds true for a current account deficit. Whether a country is in a surplus or deficit position reveals nothing about the welfare of domestic residents. A large current account deficit may, for example, may result from either a domestic recession or the anticipation of rapid growth in GDP.

The relative price of consumption across time is given by the real rate of interest. For an individual (or small open economy), one may usefully view the interest rate as exogenous. However, in the grand scheme of things, interest rates are just prices that must at some level reflect the underlying structure of the economy (e.g., preferences and technology). Taking all economies together, net financial saving must add up to zero. Thus, the interest rate can be thought of as being determined by the requirement that the sum of desired net (financial) saving is equal to zero (i.e., that the supply of credit equals the demand for credit).

4.7 Problems

1. Dominica is a small Caribbean nation (population approximately 70,000 people) whose main industry is banana production (26% of GDP and 40% of the labor force). This island nation is frequently hit by tropical storms, sometimes of hurricane strength. From Figure 4.16, we see that these storm episodes are associated with movements in GDP and net exports below their trend levels. Note as well that private consumption spending remains relatively stable throughout these episodes. How would you explain these general patterns in this data?

Figure 4.16
Dominica
Real per capita GDP and Components
1977 - 1996



2. From Figure 4.16, does it appear that the Dominican economy suffers from 'borrowing constraints?'

3. Suppose that consumer spending rises in the current quarter and that this is followed by an increase in GDP in the following quarter. Based on this observation alone, would it be safe to conclude that strong consumer

spending 'caused' the rise in future GDP? If your answer is no, explain why not. If your answer is yes, then explain: (1) what may have caused consumer spending to rise in the first place; and (2) how this increase in consumer spending led to a higher GDP.

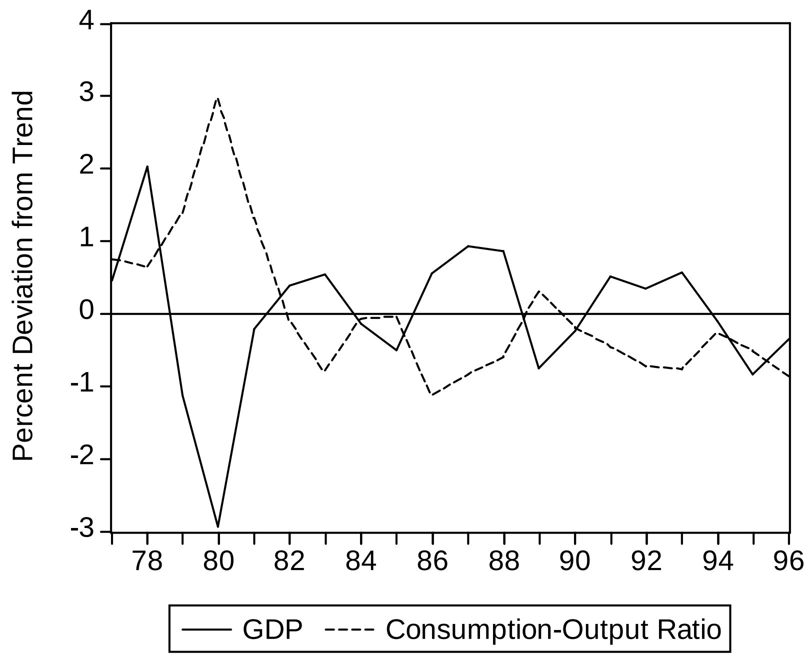
4. Consider the following quote from a recent commentary by James Arnold (BBC News): "Consumer spending is certainly the foundation of many economies. The long boom of the mid to late 1990s was built on buoyant spending - especially in the US and UK, where service industries have long replaced manufacturing as the main economic motor. Similarly, the predicted slump in consumer spending is seen as the main threat now, as the US attacks (9/11) crunched into an already-vulnerable global economy." (Note: the predicted slump in consumer spending did not materialize). The quote seems to suggest that economic growth is driven by (presumably exogenous) consumer spending. Offer a critique of this perspective.

5. Suppose that preferences are such that $MRS = c_2/c_1$. Show that the consumption-output ratio (c_1^D/y_1) is given by:

$$\frac{c_1^D}{y_1} = \frac{1}{2}\left(1 + R^{-1}\frac{y_2}{y_1}\right).$$

Explain why the consumption-output ratio is likely to be countercyclical in an economy subject to transitory productivity shocks, but relatively stable in an economy subject to permanent productivity shocks. Is the behavior of the consumption-output ratio for Dominica consistent with theory? (see Figure 4.17).

Figure 4.17
Dominca
Real per capita GDP and the
Consumption-Output Ratio



6. Explain why a country's current account position is a poor measure of economic welfare.

7. Using a diagram similar to Figure 4.12, show how the real interest rate is likely to react if the world financial market suddenly receives information that leads to an upward revision in the forecast of future GDP. Explain.

4.8 References

1. Andolfatto, David (2002). "A Theory of Inalienable Property Rights," *Journal of Political Economy*, 110(2): 382-393.

2. Bluedorn, John (2002). "Hurricanes: Capital Shocks and Intertemporal Trade Theory," Manuscript.

3. Fisher, Irving (1930). *The Theory of Interest*, New York: The Macmillan Company.

4. Friedman, Milton (1957). *A Theory of the Consumption Function*, Princeton NJ: Princeton.

4.A Alexander Hamilton on Repaying the U.S. War Debt

Source: www.clev.frb.org/Annual99/theory.htm#alex

Anyone who has ever spoken the words “just this once” has probably learned the hard way the problems of a time-inconsistent strategy. Time inconsistency refers to a situation in which what looks like the best decision from moment to moment may not produce the best outcome in the long run. That is, the long-term plans of people and governments often fall apart because people are free to make decisions that offer instant gratification at any point in time. Indeed, time inconsistency is a commonly faced problem in the establishment of economic policy.

After the American Revolution, Alexander Hamilton, as the first U.S. Secretary of Treasury, was given the task of refunding and repaying enormous war debts. In a report to Congress in 1790, the whole expense of the war was estimated to be $135 million. Of this amount, $5 million was owed to foreigners, $17 million was owed for supplies paid by certificates, $92 million was owed for wages and supplies paid for by “cash” redeemable in gold or silver, and $21 million was owed by the states. While it was widely agreed that money borrowed from foreign governments needed to be repaid, many in the new Congress, including Thomas Jefferson and James Madison, argued against the repayment of some obligations to avoid the difficulties that increased taxation would cause.

But Hamilton was committed to establishing the government’s creditworthiness. He knew the dangers of defaulting on debt, or implicitly defaulting by engineering inflation. Hamilton understood that by taking the expedient course and defaulting on some holders of the war debt, Congress would cast doubt on the trustworthiness of the new government to honor its debts. In so doing, they would inadvertently drive up the cost of credit by reducing the appeal to investors that the nation so desperately needed. In other words, his model was time consistent.

Hamilton felt so strongly about his position that he agreed to endorse a plan for moving the nation’s capital from New York to Washington, D.C., if his debt repayment plan passed in Congress. Hamilton’s plan did pass, the young nation established its creditworthiness, and to this day the seat of the U.S. government shuts down if it snows more than an inch.

4.B Milton Friedman Meets John Maynard Keynes

Many of you have likely already encountered a theory of consumption in your introductory macroeconomics class called the *Keynesian consumption function.* The Keynesian consumption function is often specified as a relationship that takes the following form:

$$C = a + bY,$$

where $a > 0$ is a parameter that denotes 'autonomous' consumer spending, and $0 < b < 1$ is a parameter called the marginal propensity to consume. This consumption function embeds the common sense notion that desired consumer spending is an increasing function of income, but that a one dollar increase in income generally results in a less than one dollar increase in consumer demand. Note that this theory makes no distinction between income changes that are perceived to be temporary or permanent.

In a debate that occurred decades ago, Milton Friedman (1957) argued that consumer demand should depend on *wealth*, not income. According to Friedman, the consumption function should be specified as:

$$C = \alpha W,$$

where $\alpha > 0$ is a parameter and W denotes wealth. Thus, according to Friedman, consumer demand should be proportional to wealth and should only depend on income to the extent that income influences wealth.

We can understand both views by appealing to our theory (which builds on the early work of Irving Fisher, 1930). In particular, suppose that preferences are such that $MRS = c_2/(\beta c_1)$. Then our theory implies a consumption function of the following form:

$$c_1^D = \left(\frac{1}{1+\beta}\right)\left[y_1 + \frac{y_2}{R}\right].$$

If we let $\alpha = 1/(1+\beta)$, then we see that our theory is consistent with Friedman's hypothesis, since $c_1^D = \alpha W$, where $W = y_1 + \frac{y_2}{R}$.

On the other hand, we can rearrange our consumption function in the following way:

$$c_1^D = \left(\frac{1}{1+\beta}\right)\left(\frac{y_2}{R}\right) + \left(\frac{1}{1+\beta}\right) y_1.$$

If we define $a = \left(\frac{1}{1+\beta}\right)\left(\frac{y_2}{R}\right)$ and $b = \left(\frac{1}{1+\beta}\right)$, then we see that our consumption function also agrees with Keynes; i.e., $c_1^D = a + by_1$.

While the two theories look similar, they can in fact have very different implications for consumer behavior. For example, consider two individuals that have the same level of wealth but different lifetime income patterns. The Friedman consumption function implies that these two individuals should have the same level of consumption, while the Keynesian consumption function implies that the person with the higher current income should have higher (current) consumer demand.

Our theory is consistent with Friedman's hypothesis when individuals are not debt constrained. But if individuals are debt constrained, then our theory supports Keynes' hypothesis. In any case, our theory is to be preferred over either because it makes explicit where the parameters a, b and α come from, as well as stating the conditions under which either hypothesis may be expected to hold.

4.C The Term Structure of Interest Rates

In reality, securities can be distinguished by (among other things) their *term to maturity*. Suppose, for example, that you wish to borrow money to purchase a home and that you plan to pay off the mortgage in ten years. There are many ways in which you might go about financing such a purchase. One strategy would be to take out a 10-year (long-term) mortgage. Such a debt instrument has a term to maturity that is equal to ten years. Alternatively, one might choose to take out a one-year (short-term) mortgage and refinance the mortgage every year for ten years. Each one-year mortgage has a term to maturity equal to one year. In practice, the interest rate you pay on a one-year mortgage will typically differ from the interest rate you would pay on a ten-year mortgage. In other words, 'short-term' interest rates typically differ from 'long-term' interest rates.

Our model can be extended so that we may distinguish between 'short' and 'long' term interest rates. To this end, assume that the economy lasts for three periods and that the endowment is given by (y_1, y_2, y_3). Here, you can interpret y_1 as the level of current real GDP; y_2 as the *current forecast* of real GDP in the 'medium' term; and y_3 as the *current forecast* of real GDP in the 'long' term. Following the logic embedded in (4.6), the real interest rate between any two adjacent periods must satisfy:

$$R_{12}^* = \frac{1}{\beta}\frac{y_2}{y_1};$$
$$R_{23}^* = \frac{1}{\beta}\frac{y_3}{y_2}.$$

These are the interest rates you would expect to pay if you were to refinance your mortgage on a period-by-period basis. In other words, the sequence $\{R_{12}^*, R_{23}^*\}$ represents a sequence of *short-term* interest rates. Notice that these short-run interest rates depend on the sequence of short-term growth forecasts in real GDP.

Using a no-abritrage condition (ask your instructor to explain this), we can compute a 'long-run' interest rate; i.e.,

$$R_{13}^* = R_{12}^* R_{23}^* = \frac{1}{\beta^2}\frac{y_3}{y_1}.$$

Here, R_{13}^* represents the total amount of interest you would pay (including principal repayment) if you were to finance your mortgage with a long-term debt instrument (i.e., if your mortgage was to come due in two years, instead of one year). Notice that the total amount of interest you would pay is the same whether you finance your mortgage on a year-by-year basis or whether you finance it with a longer-term debt obligation. The annual (i.e., geometric average) rate of interest you are implicitly paying on the longer-term mortgage is given by:

$$R_L^* = (R_{13}^*)^{1/2} = \left(\frac{1}{\beta^2}\frac{y_3}{y_1}\right)^{1/2}.$$

Notice that this 'long-run' interest rate depends on the 'long-run' forecast of real GDP growth.

The pair of interest rates $\{R_{12}^*, R_L^*\}$ (which are both expressed in annual terms) is called the *term structure of interest rates* or the *yield curve*. As of period 1, R_{12}^* is the 'short-run' interest rate (or yield) and R_L^* is the 'long-run' interest rate (or yield). The yield curve is a graph that plots these interest rates on the y-axis and the term-to-maturity on the x-axis. The difference $(R_L^* - R_{12}^*)$ is called the *slope of the yield curve*.

In reality, we observe that short-run interest rates are much more volatile that long-run interest rates. A simple explanation for this (based on our theory) is that long-run forecasts of GDP growth are relatively stable whereas forecasts of short-run growth are relatively volatile (this would be the case, for example, if there is a transitory component in the GDP growth rate).

Assuming that the long-run growth rate of GDP is relatively stable, the slope of the yield curve can be used to forecast the likelihood of recession or recovery. Suppose, for example, that we are in a recession (in the sense that y_1 has in some sense 'bottomed out.'). In this case, the slope of the yield curve is likely to be negative. The negative slope of the yield curve is signalling the market's expectation of an imminent recovery (i.e., the forecast of near-term growth (y_2/y_1) is relatively high). On the other hand, imagine that y_1 is currently near a 'normal' level and that the short-run interest suddenly drops (with long-term rates remaining relatively stable). In this case, the slope of the yield curve turns positive, signalling the market's expectation of near-term weakness in GDP growth.

4.D The Intertemporal Substitution of Labor Hypothesis

In the model studied in the this chapter, time-dated consumption goods (and services) were viewed as distinct commodities (like apples and oranges). The relative price of time-dated output is the real interest rate. Changes in the interest rate, like changes in any relative price, induce people to substitute across commodities (the substitution effect). For example, an increase in the interest rate makes current consumption more expensive relative to future consumption, inducing people to save more (delay consumption to when it is cheaper).

The basic logic of this argument applies not only to consumption, but to leisure (and therefore labor) as well. To see how this works, let us combine the model developed in Chapter 2 with the model developed above (for the case of a small open economy). In particular, assume that individuals have preferences defined over both time-dated consumption and leisure. Furthermore, suppose that these preferences take the following form:

$$u(c_1, l_1, c_2, l_2) = [\ln c_1 + \lambda \ln l_1] + \beta \ln [\ln c_2 + \lambda \ln l_2],$$

where λ and β are preference parameters. For these preferences, we can identify three relevant marginal rates of substitution:

$$MRS(c_j, l_j) = \lambda \frac{c_j}{l_j}; \text{ for } j = 1, 2;$$
$$MRS(c_1, c_2) = \frac{1}{\beta}\frac{c_2}{c_1}.$$

Following Chapter 2, assume that the production function is given by $y_j = z_j n_j$ and that $n_j + l_j = 1$ for $j = 1, 2$. Recall that in this world, the equilibrium wage rate in each period will be equal to z_j. Therefore, optimizing along the intratemporal dimension requires:

$$\lambda \frac{c_j}{1 - n_j} = z_j \text{ for } j = 1, 2. \tag{4.7}$$

From what we learned in this chapter, if individuals are free to borrow and lend at the interest rate R, then optimizing along the intertemporal dimension requires:

$$\frac{1}{\beta}\frac{c_2}{c_1} = R. \tag{4.8}$$

These choices will be constrained by an intertemporal budget constraint:

$$c_1 + \frac{c_2}{R} = z_1 n_1 + \frac{z_2 n_2}{R}. \tag{4.9}$$

Equations (4.7), (4.8) and (4.9) constitute four restrictions that must hold in equilibrium (for a given R). These four equations can be solved the four unknowns $(c_1^D, c_2^D, n_1^*, n_2^*)$. Note that the 'stars' on the employment levels indicate that these are 'general equilibrium' quantities in the sense that wages are clearing the domestic labor market in each period. Given these employment levels, the economy's real GDP flow is determined by $y_j^* = z_j n_j^*$, for $j = 1, 2$. But since this is a small open economy, there is no requirement for $c_j^D = y_j^*$ (in general, these two variables will not be equal).

Note that since R is exogenous, the equilibrium employment levels (and hence, the GDP flows) will, in general, depend on R. To see how this might work, consider equations (4.7), which may be rewritten as:

$$\lambda c_1 = z_1(1 - n_1);$$
$$\lambda c_2 = z_2(1 - n_2).$$

Now, divide the second equation by the first equation to get:

$$\frac{c_2}{c_1} = \frac{z_2(1 - n_2)}{z_1(1 - n_1)}.$$

From equation (4.8), we see that $(c_2/c_1) = R\beta$. This allows us to rewrite the equation above as:

$$R\beta = \frac{z_2(1 - n_2)}{z_1(1 - n_1)}. \tag{4.10}$$

Equation (4.10) allows us to identify the substitution effects (but not the wealth effects) that are at work in this economy.

First, consider what happens if R increases. Equation (4.10) tells us that either n_2 must fall or n_1 must rise (or some combination of both). In other words, an increase in the interest rate has the effect of increasing current employment (and output) relative to future employment (and output). The basic intuition here is as follows. The higher interest rate makes it a good time to work, since by working you can generate income with which you can use to increase your current saving (since the return to saving is higher). You can then use this extra saving in the future to enjoy higher levels of leisure (allowing you to work less in the future). This is one dimension of the intertemporal substitution of labor hypothesis. the quantitative relevance of this effect in reality appears to be small.

The second dimension of the intertemporal substitution of labor hypothesis concerns the effects of transitory versus permanent changes in real wages. A permanent change in the real wage can be modeled as $\Delta z_1 = \Delta z_2$. According to (4.10), a permanent change in the wage has no effect on *relative* labor supplies (of course, labor supply may change in all periods owing to a wealth effect). On the other hand, 4.10) also reveals that transitory changes in the wage are likely to induce an intertemporal substitution of labor. For example, consider a temporary increase in the current wage; i.e., $\Delta z_1 > 0 = \Delta z_2$. The effect of this change is for workers to increase n_1 relative to n_2. In other words, when the return to labor is perceived to be temporarily high, individuals choose to work harder today (and rest more tomorrow). This type of intertemporal substitution may play a significant role in explaining why employment varies so much across seasons.

Chapter 5

Government Spending and Finance

5.1 Introduction

In this chapter, we continue with our small open economy analysis and investigate the effects of government spending and finance. This analysis differs from that of Chapter 3 in that intertemporal dimensions are now explicitly considered. Among other things, this will allow us to think clearly about the effects of government budget deficits and surpluses. This is in contrast to the static model studied in Chapter 3 where we had to assume that the government balanced its budget on a period-by-period basis.

5.2 The Government Budget Constraint

As in Chapter 3, we do not ask why there is a government; we just assume that there is one and that it requires resources. Assume that the legislative branch of the government has in place an *expenditure program*. We can model this program as a pair (g_1, g_2). Notice that when one speaks of 'an increase in government spending,' one has to be careful to specify whether this increase is temporary, anticipated, or permanent.

Exactly how the government's expenditure program is to be financed is a problem that falls on the government's finance department (e.g., the *Department of Finance* in Canada; the *Treasury Department* in the United States; and the *Ministry of Finance* in Japan). Finance departments typically have only one way in which to secure the resources necessary to finance the government's demand for goods and services: taxes. While the finance department can also secure resources by borrowing (i.e., by issuing government bonds), to the extent that these bonds are eventually repaid, they simply constitute *future* taxes. Consequently, the problem facing a finance department basically boils down to one of choosing the appropriate *timing of taxes*.

Assume, for simplicity, that the finance department can implement lump-sum taxes (τ_1, τ_2). One finance plan that could balance the budget is then given by $\tau_1 = g_1$ and $\tau_2 = g_2$. But if a government has access to a financial market, it need not balance its budget on a period-by-period basis. In particular, the government could choose to save or borrow (just like individuals).

Recall from Chapter 4 that we defined 'saving' in general terms as current income minus expenditures on current needs. In the present context, we can define 'government saving' as:

$$s_G \equiv \tau_1 - g_1. \tag{5.1}$$

That is, if $\tau_1 > g_1$, then the government is bringing in more tax revenue than it needs to finance current spending requirements. In this case, we say that the government is running a *budget surplus*. If $\tau_1 < g_1$, then the government's tax revenues are not sufficient to meet current spending requirements. In this case, we say that the government is running a *budget deficit*. Note that the government budget deficit (b_G) can be defined as the negative of government saving; i.e., $b_G = -s_G = g_1 - \tau_1$. You can think of b_G as representing the government bonds that must be issued in order to cover the short fall in tax revenue.

In the second period, the finance department must set taxes at a level that cover both expenditures g_2 and its debt obligations Rb_G; i.e.,

$$\tau_2 = g_2 + Rb_G. \tag{5.2}$$

By substituting $b_G = g_1 - \tau_1$ into (5.2), we derive the following expression:

$$g_1 + \frac{g_2}{R} = \tau_1 + \frac{\tau_2}{R}. \tag{5.3}$$

The expression in (5.3) constitutes the government's *intertemporal budget constraint*. Notice the resemblance to the household's intertemporal budget constraint in Chapter 4. In particular, the right-hand-side of (5.3) measures the present value of the tax (revenue) stream; as such, it constitutes the government's wealth. The left-hand-side measures the present value of the government's expenditure program. The intertemporal government budget constraint does not imply that the government must balance its budget on a period-by-period basis (although, this is certainly feasible). In particular, a government may (if it wishes) run a deficit. But to do so, it must at some point in the future plan to run a surplus–at least, if it plans to make good on its debt obligations.

5.3 The Household Sector

Individuals in the household sector face the same decision problem as in Chapter 4, except that they now have a tax obligation (τ_1, τ_2) to deal with. If we define $y_j - \tau_j$ as *disposable income*, then private sector saving can be defined as:

$$s_P \equiv y_1 - \tau_1 - c_1,$$

with the second-period budget constraint is given by:

$$c_2 = y_2 - \tau_2 + Rs_P.$$

Combining these latter two expressions, we can form the individual's intertemporal budget constraint:

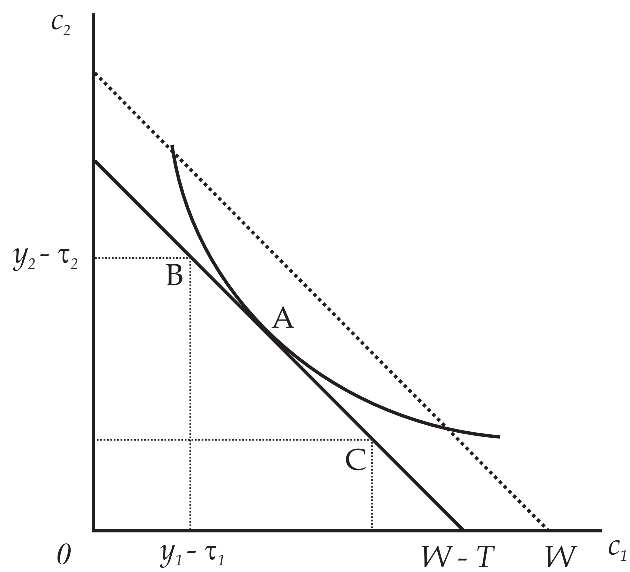
$$c_1 + \frac{c_2}{R} = (y_1 - \tau_1) + \frac{(y_2 - \tau_2)}{R}.$$

It will be convenient to rewrite the household's intertemporal budget constraint in the following manner:

$$c_1 + \frac{c_2}{R} = W - T,$$

where $W = y_1 + R^{-1}y_2$ denotes gross wealth and $T = \tau_1 + R^{-1}\tau_2$ denotes the present value of a household's tax obligation to the government. Hence, the term $W - T$ represents the household's after-tax wealth. The solution to the household's choice problem is depicted in Figure 7.1 (point A).

FIGURE 5.1
Individual Choice with Lump-Sum Taxes



There is something important to observe here. If individuals live in a 'Friedman' world (see Appendix 4.B), then they *do not care about the timing of their tax payments, if changes in the timing result in the same lifetime tax obligation* T. For example, Figure 5.1 depicts two after-tax endowment profiles that result in the same after-tax wealth; endowment B features high current taxes (but low future taxes), while endowment C features low current taxes (but high future taxes). In either case, consumer demand remains at point C. On the other hand, if individuals live in a 'Keynesian' world (i.e., if they are debt constrained), then the same conclusion will generally not hold (again, see Appendix 4.B). The importance of this distinction will become apparent shortly. In the meantime, we will operate under the assumption that individuals are not debt constrained.

5.4 The Ricardian Equivalence Theorem

In this section, we ask two related questions. First, how does a cut in taxes affect consumer demand? Second, does a large government budget deficit pose any sort of 'problem' for the economy? These two questions are related because cutting taxes generally implies increasing the deficit, at least, to the extent that program spending (g_1, g_2) is left unaltered. Another way to ask the question being posed here is: What are the likely effects of a deficit-financed tax cut?

Deficit-financed tax cuts are sometimes recommended by policy advisors when the economy is in recession. The reasoning here runs something as follows. First, we know that increases in consumer demand are often followed by periods of economic expansion. If consumer spending is an increasing function of disposable income (e.g., $c = a + b(y - \tau)$ as in Appendix 4.B), then a cut in taxes will increase the disposable income of the household sector, leading to an increase in consumer demand and therefore future GDP.

Let us investigate the logic of this argument within the context of our model. Take a look at the government's budget constraint (5.3). If we hold the pattern of government spending (g_1, g_2) fixed, then a tax-cut today $\Delta\tau_1 < 0$ *must imply a future tax increase.* This is because the deficit incurred today must be repaid (principal and interest) at some point in the future. The government budget constraint makes it clear that future taxes must rise by the amount $\Delta\tau_2 = -\Delta\tau_1 R > 0$.

The key question here is how the deficit-financed tax cut affects the after-tax wealth of the household sector. Since gross wealth W is fixed by assumption, after-tax wealth can only change if the present value of the household sector's tax liability T changes. The change in the tax liability is given by:

$$\Delta T = \Delta\tau_1 + \frac{\Delta\tau_2}{R}.$$

Observe that since $\Delta\tau_2 = -\Delta\tau_1 R$, it follows that $\Delta T = 0$.

Because the deficit-financed tax cut leaves the after-tax wealth position of the household unchanged, we can conclude that this program will have absolutely *no effect* on aggregate consumer demand. Another way to state this result is to assert that 'deficits do not matter.' The intuition behind this result is straightforward. While the current tax cut increases current disposable income of our model households, these households are also forecasting a future tax hike and hence a reduction in their future disposable income. The consumption smoothing motive tells us that households would want to react to such a change in the intertemporal pattern of their disposable income by increasing their current desired saving. By doing so, they can shift the current tax windfall to the future, where they can use it to pay for the higher taxes in that period. Since after-tax wealth is left unchanged, households increase their desired saving dollar-for-dollar with the decrease in public sector saving; i.e., $\Delta s_P = -\Delta s_G = \Delta b_G$. In other words, all the new bonds that are issued by the government are willing purchased by the household sector at the prevailing interest rate, leaving desired national saving unchanged. When these bonds mature in the future, they are used by households to pay off the higher tax bill.

The conclusion that 'deficits do not matter' is a result implied the *Ricardian Equivalence Theorem.* Loosely speaking, the Ricardian Equivalence Theorem asserts that under some conditions (that we will talk about shortly), taxes and deficits are *equivalent* ways of financing any given government expenditure stream. That is, since deficits simply constitute future taxes, the theorem alternatively asserts that *the timing of taxes do not matter.* Another way of stating the same thing is that the household sector should not view its government bond holdings as net wealth since such bonds simply represent a future tax obligation (Barro, 1974).[1]

- **Exercise 5.1.** If the Ricardian Equivalence Theorem holds, then the timing of taxes 'do not matter' in the sense that there is no effect on consumer demand, desired national saving, the current account and (in a closed economy) the real rate of interest. However, the timing of taxes does have implications for the composition of desired national saving (between the private and public sectors). Explain how.

- **Exercise 5.2.** True, False or Uncertain and Explain. The Ricardian Equivalence Theorem states that government spending 'does not matter.' (Hint: the answer is False).

The conclusions of the Ricardian Equivalence Theorem are both striking and controversial, so let us take some time now to examine the assumptions underlying these results. The theorem makes an number of important assumptions (that happen to hold true in our model economy). These assumptions are stated below:

1. *Perfect financial markets.* That is, individuals are free to save and borrow at the market interest rate. In particular, if some individuals are debt-constrained, then the theorem does not hold. On the other hand, if only a small number of people are debt-constrained, then the assumption of perfect financial markets might serve as a reasonably good approximation.

2. *'Rational' households.* In particular, households must be 'forward looking' and understand the government budget constraint. While it is easy to imagine that there may be 'irrational' households operating in the real world, one would have to question whether these households influence aggregate expenditure in a quantitatively important way. It is equally apparent by the fact that households save that they are forward looking. And judging by the political controversy generated by budget deficits, it seems hard to believe that households are generally not aware of the government budget constraint.

3. *Lump sum taxes.* In particular, the theorem does not hold if the government only has access to distortionary taxes. Since distortionary taxes are the norm in reality, this assumption is potentially a serious one.

4. *Long-lived households.* What we literally need here is that the planning horizon of the household is as long as the government's planning horizon. Since governments typically live much longer than individuals, one might question the empirical relevance of this assumption. To see what can 'go wrong' if households have short planning horizons, consider the case of

[1] See also: www.garfield.library.upenn.edu/ classics1992/ A1992GX22600001.pdf

an individual in retirement. If the government cuts this person's taxes today and increases taxes at some point in the distant horizon, then our retired individual is unlikely to 'be around' to settle the higher future tax bill (he will have cleverly escaped his tax obligation by dying). For such an individual, a deficit-financed tax cut constitutes an increase in wealth. On the other hand, while individuals do not live forever, it is conceivable that households do. Barro (1974) has pointed out that to the extent that people care about their children, they may want to save the tax cut and bequest it to their children (who can then use it to pay for the higher taxes they will face).

The Ricardian Equivalence Theorem clearly makes some strong assumptions, most of which are literally not true in reality. However, whether an assumption is literally true or not is not the relevant issue. The relevant question is whether the set of assumptions serve as good approximations to reality. Whether a set of assumptions serve as good approximations or not can only be judged by subjecting the theory to empirical testing.

As it turns out, empirical tests of the Ricardian Equivalence Theorem report are mixed (try performing a search on *Google*). Many empirical studies find that an increase in budget deficits (a decrease in public sector saving) is met by an increase in private sector saving, as the theorem predicts. However, it is less clear whether private savings rise dollar for dollar with the decline in government saving (as the theorem also predicts).

Perhaps the main lesson of the theorem for policy makers is as follows. To the extent that households increase their saving in response to a deficit-financed tax cut, such a policy is not likely to be as stimulative as one might expect (if one was trained to view the world through the lens of the Keynesian consumption function).

- **Exercise 5.3.** Explain why the Ricardian Equivalence Theorem is unlikely to hold in an economy that experiences net immigration flows.

- **Exercise 5.4.** Demonstrate, with the aid of a diagram, how the Ricardian Equivalence Theorem will not hold for an economy where individuals are debt-constrained.

5.5 Government Spending

It is important to understand that while our model implies that government budget deficits 'do not matter,' the same is not true of government spending. In our model, changes in the government expenditure program (g_1, g_2) will matter, at least, to the extent that it alters the after-tax wealth position of the household sector.

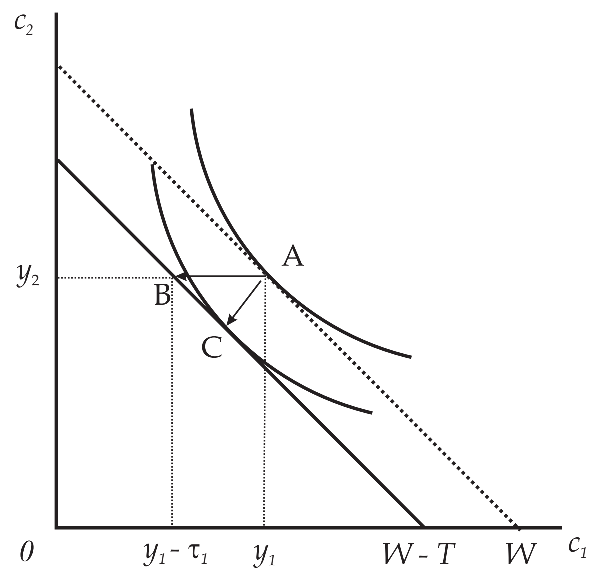
The results of this section can be summarized briefly. First, since we are working with an endowment economy, changes in (g_1, g_2) can have no effect on real output (y_1, y_2). Any increase in government spending then must ultimately imply lower levels of private consumer spending. Second, since the Ricardian Equivalence Theorem holds in our model, we can without loss of generality assume that $\tau_1 = g_1$ and $\tau_2 = g_2$. That is, since the timing of taxes 'does not matter,' let's just assume that the government balances its budget on a period by period basis. In this case, domestic saving corresponds to private sector saving (since public sector saving will always be equal to zero).

5.5.1 A Transitory Increase in Government Spending

Consider an initial situation in which $(g_1, g_2) = (0, 0)$ and suppose that households are initially content with consuming their endowment; i.e., point A in Figure 5.2 (remember that where you place the initial indifference curve does not matter). A transitory increase in government spending can be modeled as $\Delta g_1 > 0$ and $\Delta g_2 = 0$. We are assuming here that $\Delta\tau_1 = \Delta g_1$, but remember that whether the government finances this increase with higher current taxes or a deficit (higher future taxes) will not matter.

This fiscal policy shifts the after-tax endowment point to the left (i.e., to point B). The higher tax burden makes households less wealthy. The consumption smoothing motive (i.e., the wealth effect) implies that generally speaking, households will react to this fiscal policy by reducing their demand for consumption at all dates; i.e., $\Delta c_1^D < 0$ and $\Delta c_2^D < 0$. We can depict this change in behavior by moving the indifference curve from point A to point C in Figure 5.2.

FIGURE 5.2
A Transitory Increase in Government Spending



From Figure 5.2, we see that current consumer spending does not decline by the full amount of the tax increase. Therefore, private sector (and domestic) saving must decline. Households react to the transitory increase in spending (and taxes) by increasing the amount they wish to borrow from foreigners. By (temporarily) increasing the net imports of goods and services, domestic consumers can smooth their consumption over time. Of course, the resulting current account deficit must be matched in the future by a corresponding current account surplus (domestic households must export goods and services to the foreign sector to pay back their debt).

- **Exercise 5.5.** Demonstrate, with the aid of a diagram, the effects of a transitory increase in government spending financed by a deficit.
- **Exercise 5.6.** Demonstrate, with the aid of a diagram similar to Figure 5.2, what effect an anticipated increase in future government spending will have on the current account.

5.6 Government Spending and Taxation in a Model with Production

The analysis above has assumed that the intertemporal production of output (y_1, y_2) is exogenous. We can move a step closer to reality by assuming instead that the level of production depends on the time-allocation choices made in the labor market, the way we described in Chapter 2 and Appendix 4.D.

In a two-period model, the preferences of households must be modified to include time-dated leisure; i.e., $u(c_1, l_1, c_2, l_2)$. If the production function is linear in labor; i.e., $y_j = z_j n_j$ for $j = 1, 2$, then using the arguments developed in Chapter 2, we know that the equilibrium gross wages in this model economy will be given by $(w_1^*, w_2^*) = (z_1, z_2)$. The household's intertemporal budget constraint then depends on whether taxes are lump sum or distortionary. For lump-sum taxes, the budget constraint is given by:

$$c_1 + \frac{c_2}{R} = z_1(1 - l_1) - \tau_1 + \frac{z_2(1 - l_2) - \tau_2}{R},$$

and the government budget constraint takes the earlier form:

$$g_1 + \frac{g_2}{R} = \tau_1 + \frac{\tau_2}{R}.$$

If taxes are distortionary, as in a tax on labor earnings, then the budget constraint is given by:

$$c_1 + \frac{c_2}{R} = (1 - \tau_1)z_1(1 - l_1) + \frac{(1 - \tau_2)z_2(1 - l_2)}{R},$$

and the government budget constraint is given by:

$$g_1 + \frac{g_2}{R} = \tau_1 z_1(1 - l_1) + \frac{\tau_2 z_2(1 - l_2)}{R}.$$

When taxes are distortionary, we see that taxes will affect the real return to labor, so that the effect on labor supply will be affected in much the same way as it would in response to a change in productivity (z_1, z_2); again, see Appendix 4.D.

5.6.1 Ricardian Equivalence

When taxes are lump sum, the Ricardian Equivalence Theorem continues to hold in this environment. However, this will not be the case if taxes are distortionary. To see why, consider what happens if the government decides to implement a deficit-financed tax cut. In this case, the tax cut today $(\Delta\tau_1 < 0)$ stimulates employment (and hence, output) today so that $\Delta n_1^* > 0$ and $\Delta y_1^* > 0$. The tax increase expected in the future $(\Delta\tau_2 > 0)$ has the opposite effect, so that $\Delta n_2^* < 0$ and $\Delta y_2^* < 0$. Clearly, the timing of taxes does matter here. We can also see why a large deficit today may elicit some concern on the part of the population. That is, if people understand that a high deficit today must at some point be met with higher future taxes, and if these taxes are distortionary, then people will understand that high deficits today will put a drag on future economic activity.

5.6.2 Government Spending Shocks

When taxes are lump sum, any type of positive government spending shock will simply serve to reduce the after-tax wealth of the household sector. When wealth declines, the demand for all normal goods declines so that $\Delta c_j^* < 0$ and $\Delta n_j^* > 0$. As in Chapter 3, a positive government spending shock (whether transitory, anticipated, or permanent), will induce an economic boom, $\Delta y_j^* > 0$ for $j = 1, 2$. Recall, however, that since private consumption and leisure decline, the increase in output will not necessarily be associated with an improvement in economic welfare.

When taxes are distortionary, individuals are hit by a 'double-whammy,' so to speak. Since higher levels of government spending require higher taxes at some point, not only do households experience a decline in wealth, but their decisions become distorted (in an attempt to escape the tax). Since these higher taxes are distortionary, they may very well lead to a decline in employment and output (again, see Chapter 3). It is for these reasons that 'supply side' economists are critical of large government spending programs.

5.6.3 Barro's Tax-Smoothing Argument

Suppose that the government's expenditure program (g_1, g_2) is fixed in place. When taxes are lump-sum, the government's finance department faces a trivial decision: choose any (τ_1, τ_2) that satisfies the government's intertemporal budget constraint. However, when taxes are distortionary, Barro (1979) has pointed out that it would be optimal for the government to smooth taxes over time. That is, the government should choose a tax rate that balances not only the government's intertemporal budget constraint, but balances government spending and revenue on average throughout time. This implies a relatively constant tax rate and a budget deficit/surplus that fluctuates over time (but balance out over the long-run).

By smoothing taxes in this manner, the government is in effect smoothing out (and therefore minimizing) the distortions that its taxes create over time. For example, if the government requires an extraordinarily high (but transitory) level of government purchases in one period (say, to finance a war effort), the tax smoothing argument implies that the government should finance such an expenditure by issuing bonds rather than by raising taxes to extraordinarily high levels. The tax rate should be increased slightly (to minimize distortions) and kept at this higher level until the debt is paid off.

5.7 U.S. Fiscal Policy

There has been much talk recently of George W. Bush's fiscal policy. In a nutshell, this policy appears to entail: (1) tax cuts (in order to stimulate economic activity); (2) an increase in government spending on the military (to fight the war on terror); and (3) a decrease in government spending in other areas. I will not attempt a full analysis of this fiscal program, but will provide some perspective in the context of the historical pattern of U.S. government spending and taxation. Figure 6.3 (should be 5.3) plots U.S. government spending and taxation (as a ratio of GDP) beginning in 1930.

Figure 6.3
U.S. Government Spending and Taxation
1930 - 2008

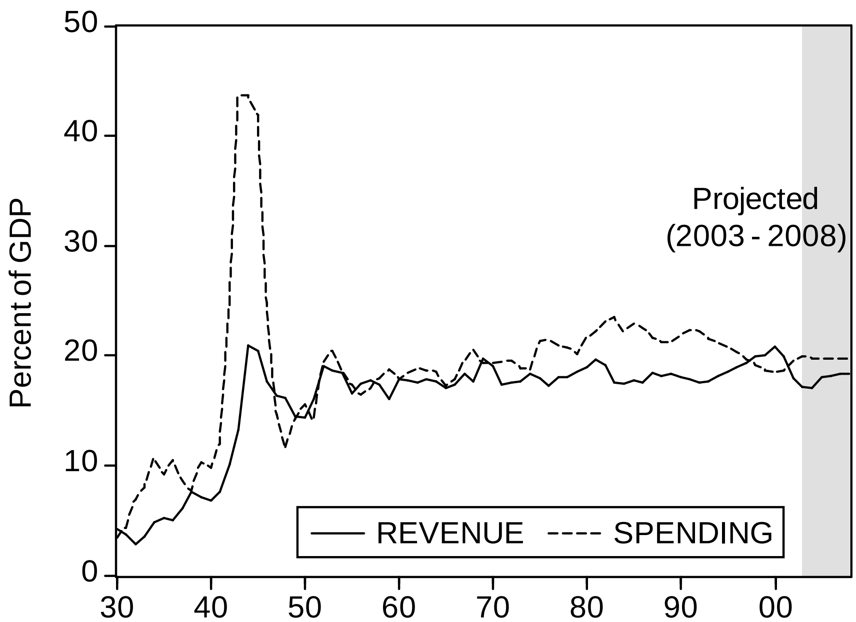


Figure 6.3 (should be 5.3) reveals a number of interesting patterns. First, note that since the end of the second world war, government spending as a ratio of GDP has remained relatively constant, displaying a moderate rise through the Carter-Reagan era, and a moderate decrease through the Clinton era. While I have not plotted it here, one should keep in mind that there has been a secular decline in the proportion of government spending devoted to the military since the end of the Korean war (in 1953, military spending was 15% of GDP; in 2000, military spending was only 4% of GDP). If one is to believe the projections in the figure, then the Bush proposal for expanded government spending plan (and associated deficits) pales in comparison to the historical data.

Also note the sharp rise in government spending during the second world war (most of which was in the form of military spending). While taxes did rise significantly during the war, they did not rise anywhere near to the extent needed to balance the budget. Here, we see Barro's tax-smoothing argument at work. That is, to the extent that the war was perceived to be transitory, it made sense to finance the bulk of expenditures by issuing bonds, rather than by raising taxes.

5.8 Summary

The intertemporal approach to government spending and finance emphasizes the fact that a government with access to financial markets is subject to an intertemporal budget constraint. From this perspective, it is clear that current budget deficits simply represent future taxes. The intertemporal approach also makes clear the importance of evaluating fiscal policy as an entire program that dictates not only current spending and taxation, but the entire future path of spending and taxation.

In some circumstances, it was shown that for a given expenditure program, the timing of taxes is irrelevant as long as the government has access to a lump sum tax instrument. This conclusion, however, is unlikely to hold empirically because taxes are typically distortionary. When taxes are distortionary, it makes sense to smooth taxes over time and allow budget deficits to grow during recessions (or periods when government spending requirements are high), followed by budget surpluses during periods of economic expansion (or periods when government spending requirements are low).

In the models studied above, government spending has the effect of 'crowding out' private consumption expenditures. Certain types of government spending shocks were also shown to affect the current account position of a small open economy. In addition to these effects, government spending is often asserted to crowd out private investment spending and lead to higher interest rates. These issues can be explored in later chapters once we have an appropriate theory of investment developed.

5.9 Problems

1. Consider a small open economy as in Figure 5.2. In that figure, we assumed that the transitory increase in government spending was financed by an increase in current taxes. Suppose instead that the government chooses to finance the current increase in government spending with an increase in *future* taxes. Show that the method of finance has no effect on desired consumer spending or the current account, but serves simply to alter the composition of national saving.

2. Consider a *closed* economy with individuals who have preferences given by $MRS = c_2/c_1$. Show that the equilibrium real interest rate is given by:

$$R^* = \frac{y_2 - g_2}{y_1 - g_1}.$$

   How is the interest rate predicted to react to: (a) a transitory increase in government purchases; and (b) an anticipated increase in future government purchases? Explain.

3. Consider an economy populated by two types of individuals, A and B. Normalize the total population to unity and let θ denote the fraction of type A individuals. Type A individuals live for one period only; their preferences are given by $u_A(c_1) = c_1$ and they have an endowment y_1. Type B individuals live for two periods; their preferences are given by $u_B(c_1, c_2) = \ln c_1 + \beta \ln c_2$ and they have an endowment (y_1, y_2). The real rate of interest is fixed at R. Imagine that a government decides to implement a public pension plan. The government plans to run this program as follows. In period one, it taxes *all* individuals an amount τ and then saves these 'contributions' at the interest rate R. In period two, the government pays out the proceeds $R\tau$ to all (living) individuals. Each person living in period two receives a payout equal to $s = R\tau/(1-\theta)$. If $g_1 = g_2 = 0$, then the government's intertemporal budget constraint is given by:

$$(1-\theta)\frac{s}{R} = \tau.$$

   The left hand side of the GBC represents the present value of the government's pension liabilities (promises). The right hand side represents the taxes that are collected in order to cover these liabilities.

   (a) Assume for the moment that $\theta = 0$. Explain why the government pension program has no effect on aggregate consumer demand. Hint: use the Ricardian Equivalence Theorem.

   (b) Now, assume that $\theta > 0$. Show that the aggregate demand for consumption in period one is now increasing in the 'generosity' of the promised payout s. Explain. Why does the Ricardian Equivalence Theorem not hold here?

5.10 References

1. Barro, Robert J. (1974). "Are Government Bonds Net Wealth?" *Journal of Political Economy*, 82: 1095–1117.

2. Barro, Robert J. (1989). "On the Determination of the Public Debt," *Journal of Political Economy*, 64: 93–110.

Chapter 6

Capital and Investment

6.1 Introduction

The model economies that we have studied so far have abstracted from *physical* capital and *investment* (expenditures on new capital goods). The models developed in Chapters 4 and 5 did feature savings, but these savings took the form of purchases (or sales) of *financial* capital. Financial capital simply represents claims against the output of other members in society (e.g., claims against the output of individuals, foreigners, or governments). However, in these previous chapters we maintained an important assumption; namely, that output is non-storable. The way to think of physical (as opposed to financial) capital is that it represents an intertemporal production technology that allows the economy to 'transport' output across time. The most obvious example of physical capital is inventory. But physical capital can also take the form of a factor of production (like labor).

Most of the physical capital in an economy constitutes durable assets that produce services that are useful in the production of output (new goods and services). Examples of such capital include the residential capital stock (which produces shelter services) and various forms of business capital (office towers, land, machinery and equipment, inventory, etc.). In most production processes, both labor and capital are important inputs for the creation of goods and services. The goods and services that are produced by these factors of production can be classified into two broad categories: consumer goods and investment goods. Investment goods are treated as expenditures on new capital goods (and include additions to inventory). These goods are produced in order to augment the existing capital stock. When the capital stock increases, more output can be produced with any given amount of labor. In this way, an economy may be able to grow even in the absence of technological progress.

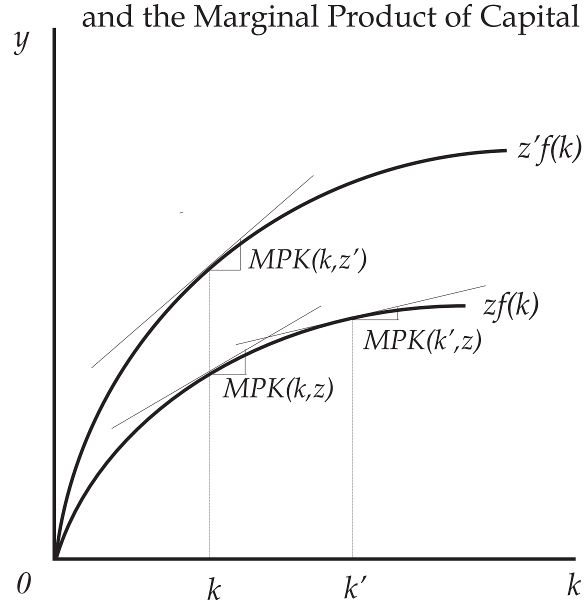
6.2 Capital and Intertemporal Production

The production function in this model economy takes the form:

$$y_j = z_j F(k_j, n_j),$$

for $j = 1, 2$. The function F is increasing and strictly concave in both k_j and n_j (see Appendix 2.A). For simplicity, let us assume that the time allocation choice is fixed (exogenous) at $(n_1^*, n_2^*) = (1, 1)$. Hence, the production function can be written as $zf(k) \equiv zF(k, 1)$, where the function f is increasing and strictly concave. All this means is that as we increase the amount of capital used in production (holding fixed the labor input), the amount of output that is produced increases. The fact that f is concave means that output increases with k at a diminishing rate (there are diminishing returns to capital accumulation). The slope of the production function is called the *marginal product of capital*, which we denote here by $MPK(k, z) \equiv zf'(k)$. The $MPK(k, z)$ tells us the extra output that can be produced by increasing capital by a small amount from k (given the technology parameter z). If the function f is increasing and concave, then it follows that the MPK is positive and a decreasing function of k. As well, note that the MPK is an increasing function of z. Figure 6.1 depicts the production technology for two different levels of technology $z' > z$. Note that $MPK(k, z') > MPK(k, z) > MPK(k', z)$.

FIGURE 6.1
Production Technology
and the Marginal Product of Capital



We assume that the economy begins period one with some exogenous amount of capital $k_1 > 0$. Think of this capital as having been determined by historical investment decisions. Since k_1 and z_1 are exogenous, and since we are assuming that $n_1^* = 1$ is exogenous, it follows that the current period GDP is exogenous as well; i.e., $y_1 = z_1 F(k_1, 1) = z_1 f(k_1)$.

The level of production at any date is imagined to consist of both new consumer goods and new capital goods. We will assume that the division of output between consumer and capital goods can be made within a period. The way to think about this assumption is that it takes firms relatively little time to reallocate factors of production across different branches of production. This is not a very realistic assumption to make if the time period is relatively short, but we make it here primarily for simplicity. The production of new capital goods serves to augment the future capital stock. Since the future level of output depends in part on the future capital stock, diverting resources away from consumption today (toward investment) serves to increase the future productive capacity of the economy.

The link between the future domestic capital stock and current domestic investment spending is given by the following identity:

$$\text{Future Capital Stock} = \text{Current Capital Stock} + \text{Net Domestic Investment}, \tag{6.1}$$

where *net domestic investment* is defined as *gross domestic investment less depreciation*. In what follows, I assume (for simplicity) that capital does not depreciate so that there is no difference between gross and net investment. I also assume that existing capital can be consumed (e.g., liquidated) after it is used in production, so that (gross and net) investment spending may conceivably be negative.

Let x_j denote expenditure on new capital goods in period $j = 1, 2$. Then utilizing the definition in (6.1), the capital stock in our model economy evolves over time according to:

$$\begin{aligned} &k_1 > 0 \text{ given;} \\ &k_2 = k_1 + x_1; \\ &k_3 = k_2 + x_2; \end{aligned} \tag{6.2}$$

Since this is a two-period model, carrying capital into the third period makes no sense (if it can be avoided), so that setting $k_3 = 0$ is desirable here. Setting $k_3 = 0$ implies setting $x_2 = -k_2 < 0$, which implies that the entire period two capital stock will be (liquidated and) consumed.

6.3 Robinson Crusoe

To begin developing ideas, it is useful to return to our Robinson Crusoe economy. Remember that Robinson Crusoe is a metaphor that describes a situation in which an individual (or an economy) cannot trade with other individuals (or economies). In the present context, this assumption rules out the existence of financial assets (which constitute claims against the output of other individuals or economies).

Crusoe has preferences defined over time-dated consumption (c_1, c_2), that for present purposes, we can interpret as 'coconuts.' Coconuts are produced with two factors of production: coconut trees (capital) and Crusoe's labor (which we have fixed at unity). In the current period, Crusoe is endowed with k_1 coconut trees (which can be liquidated and consumed). These trees (together with Crusoe's labor) produce $y_1 = z_1 f(k_1)$ units of new coconuts.

The supply of coconuts can be allocated to one of two uses: consumption or investment. Investment here means planting coconuts in the ground as seed crop if $x_1 > 0$ and liquidating existing coconut trees if $x < 0$. Assume that coconuts cannot be stored across time as inventory (although, of course, existing coconut trees survive into the next period). The act of planting coconuts in the ground constitutes an act of saving, which in this case takes the form of domestic investment in the island's coconut plantations. We thus have the following relation:

$$c_1 + x_1 = y_1. \tag{6.3}$$

Notice that the income-expenditure identity $Y \equiv C + I$ is satisfied for this economy in the first period. Note as well from the definition of saving $s \equiv y_1 - c_1$, that all saving in this economy will take the form of domestic capital expenditure. From equation (6.3), we can calculate the maximum first period consumption as $c_1^{\max} = y_1 + k_1$, which can be achieved by liquidating the entire current stock of capital; i.e., $x_1 = -k_1$. Likewise, the maximum level of investment can be attained (the minimum amount of consumption) by investing the entire GDP; i.e., $x_1^{\max} = y_1$ (so that $c_1^{\min} = 0$).

From (6.2), we see that the level of investment today affects the future stock of productive capital; i.e., $k_2 = k_1 + x_1$. This future stock of coconut trees (together with Crusoe's future labor) produces $y_2 = z_2 f(k_1 + x_1)$ units of future coconuts. Since the economy ends at the end of period two and since existing coconut trees can be consumed, Crusoe simply consumes everything he can in the second period; i.e., $c_2 = y_2 + k_2$ (again, note that the income-expenditure $C + I = Y$ identity holds, with $I = -k_2$).

Since $c_2 = y_2 + k_2 = z_2 f(k_1 + x) + (k_1 + x_1)$, we can use (6.3) to write:

$$c_2 = z_2 f(k_1 + y_1 - c_1) + (k_1 + y_1 - c_1). \tag{6.4}$$

This equation is called the *production possibilities frontier* (PPF). The PPF defines the combinations of (c_1, c_2) that are technologically feasible for Crusoe, given the parameters y_1, k_1, z_2, the structure of f, and assuming that the existing capital stock has a liquidation value equal to k_1. In particular, note that the maximum feasible level of future consumption is achieved by setting $c_1 = c_1^{\min} = 0$; i.e., $c_2^{\max} = z_2 f(k_1 + y_1) + (k_1 + y_1)$. Likewise, the maximum feasible level of current consumption is achieved by setting $c_2 = c_2^{\min} = 0$.

The slope of the PPF is given by:

$$\frac{\Delta c_2}{\Delta c_1} = -z_2 f'(k_1 + y_1 - c_1) - 1.$$

The marginal rate of transformation (between c_1 and c_2) is defined as that absolute value of the slope of the PPF; i.e.,

$$MRT(c_1, z_2) \equiv 1 + z_2 f'(k_1 + y_1 - c_1) = 1 + MPK(k_2, z_2).$$

The MRT tells us how many additional units of c_2 can be attained by sacrificing one unit of c_1 (i.e., diverting current output away from consumption and into investment). By increasing current investment by one unit, the future capital stock is increased by one unit. This extra unit of future capital can be consumed after it yields the added return MPK, so that the extra total future output available from the one unit of current investment is given by $1 + MPK$. For this reason, you can think of the (expected) future MPK as representing the (net) rate of return on current investment spending. Figure 6.2 displays the PPF in a diagram.

FIGURE 6.2
Intertemporal Production
Possibilities Frontier

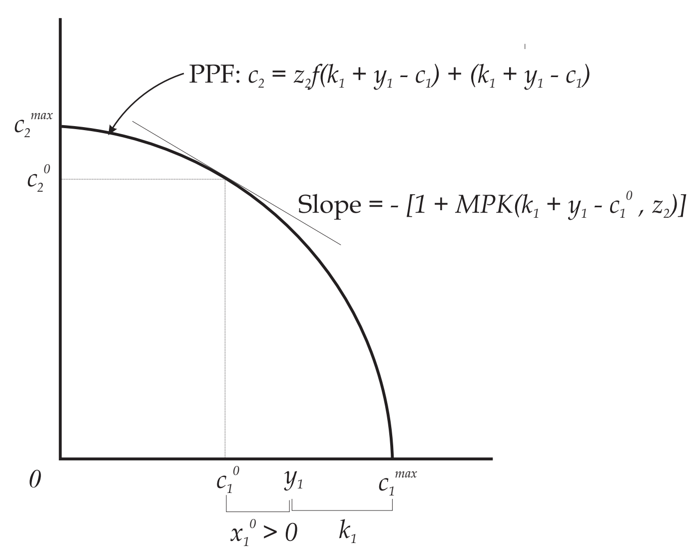
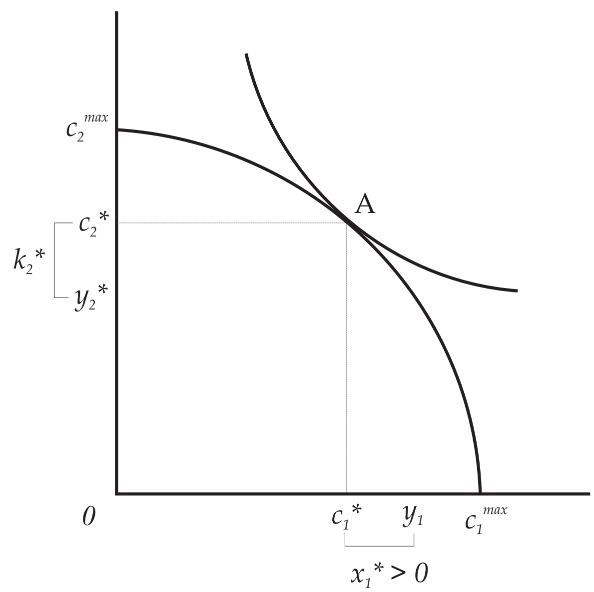


Figure 6.2 makes it clear that the availability of capital and investment allows Crusoe to choose an intertemporal pattern of production of consumer goods and services. In the endowment economy studied in Chapter 4, the intertemporal production of consumption was fixed (as an endowment). The availability of capital k_1 implies that Crusoe is now able to consume more than the current GDP; i.e., $c_1 > y_1$ is feasible (by drawing down the existing stock of capital). As well, the availability of an investment technology implies that Crusoe may alter the future level of GDP (and future consumption) by allocating resources away from current consumption toward investment.

Figure 6.2 describes what is *possible* for Crusoe to attain. But in order to understand how he actually behaves, we have to consider his preferences for time dated consumption. Crusoe's choice problem is to choose the intertemporal pattern of consumption that maximizes his well-being as measured by $u(c_1, c_2)$ subject to his constraints. The solution to this choice problem is displayed in Figure 6.3 as point A.

FIGURE 6.3
Robinson Crusoe



The solution at point A is characterized by the following two mathematical conditions:

$$MRS(c_1^*, c_2^*) = 1 + MPK(k_1 + y_1 - c_1^*); \tag{6.5}$$
$$c_2^* = z_2 f(k_1 + y_1 - c_1^*) + (k_1 + y_1 - c_1^*).$$

The first condition says that the slope of the indifference curve must be equal to the slope of the PPF; the second condition says that the solution must lie on the PPF. Only point A in the commodity space satisfies these two conditions simultaneously. Once c_1^* is known, the optimal level of (first period) investment can be easily calculated as $x_1^* = k_1 + y_1 - c_1^*$.

6.4 A Small Open Economy

Now that we understand how a Robinson Crusoe economy operates when investment is possible, let us examine how things change when we open up the economy to international trade. We alter the environment by supposing that Crusoe now has access (say, through the internet) to a world financial market which allows him to borrow or save at the world (and hence domestic) interest rate R. We treat R as an exogenous variable.

We can organize our thinking by breaking up Crusoe's choice problem into two stages. In stage 1, Crusoe chooses a level of domestic investment spending to maximize his wealth, subject to production possibilities. In stage 2, Crusoe chooses his consumption pattern (and hence the net level of international borrowing or lending) in order to maximize utility subject to his intertemporal budget constraint.

6.4.1 Stage 1: Maximizing Wealth

Imagine that you are Crusoe and that you are holding a coconut in your hand. You are thinking about saving this coconut. What are your choices? You could plant the coconut in the ground (i.e., invest it domestically). Doing so will increase your future coconut capital by $\Delta k_2 = 1$ unit. This extra future capital will then yield an additional $1 + z_2 f(k_2 + 1)$ coconuts in the future. As it turns out, this additional output is (approximately) equal to $1 + MPK(k_2, z_2)$, which is why we think of the MPK the 'rate of return on domestic capital investment.'

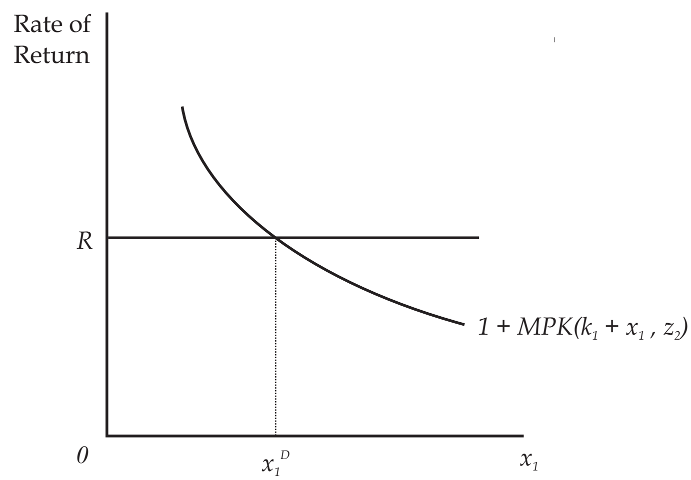
Alternatively, you could purchase a foreign bond that yields the (gross) interest rate R (or net interest rate $r = R - 1$).The way that you purchase this bond is by exporting your coconut to a foreigner in exchange for his promise to deliver $R = (1 + r)$ coconuts to you in the future. Clearly, if $MPK > r$, then you should invest the coconut domestically. If $MPK < r$, then you should export the coconut in exchange for foreign bonds.

Now since the MPK depends on the level of domestic investment spending, it would make sense to adjust investment (the future capital stock) to a point where:

$$MPK(k_1 + x_1^D, z_2) = r. \tag{6.6}$$

The way to convince yourself that this must be true, suppose that the condition does not hold. If $MPK > r$, then you could profit by borrowing a coconut from foreigners at interest rate R and planting the coconut domestically. Your net profit from such a transaction would be $MPK - r > 0$. Consequently, in order to profit in this way, you would increase domestic capital spending until no such profit opportunity remained. Remember that since the MPK falls as k_2 expands, eventually MPK will equal r as k_2 (x_1) is increased. Alternatively, if $MPK < r$, then you could profit by scaling back on domestic investment and lending the resources to foreigners at the higher return r. Your net profit from such a transaction would be $r - MPK > 0$. Of course, as you continue to scale back on domestic investment spending, the MPK rises and eventually equals r. Therefore, the desired level of domestic investment spending must equal (6.6) if no such profit opportunities are to be left unexploited. Equation (6.6) determines the investment demand function x_1^D (as a function of R, k_1 and z_2); see Figure 6.4.

FIGURE 6.4
Determination of Domestic
Investment Demand



- **Exercise 6.1.** Using Figure 6.4, show that the investment demand function $x_1^D(R, k_1, z_2)$ is a decreasing function of R and k_1, and an increasing function of z_2. Explain.

We now move to show formally how the choice of x_1^D maximizes Crusoe's wealth. To do this, let us think of Crusoe in his role as a business manager. As of period one, the business owns k_1 units of capital (measured in units of current output). This capital is used to produce the output $y_1 = z_1 f(k_1)$. If the business spends x_1 units of output on investment, then the period 'cash flow' is given by $c_1 = y_1 - x_1$. It is useful here to think of c_1 as representing the supply of consumer goods and services in period one. Note that this supply may exceed

y_1 if, for example, Crusoe chose to liquidate (transform into consumer goods) some of his current capital k_1 after production.

The investment x_1 determines the future capital stock k_2 and hence, the future level of GDP, $y_2 = z_2 f(k_2)$. The future 'cash flow' (the supply of future consumption) is therefore given by $c_2 = k_2 + z_2 f(k_2)$, since k_2 will at this point be liquidated into consumer goods. Since $k_2 = k_1 + x_1 = k_1 + y_1 - c_1$, we can alternatively write the supply of future consumption as:

$$c_2 = (k_1 + y_1 - c_1) + z_2 f(k_1 + y_1 - c_1). \tag{6.7}$$

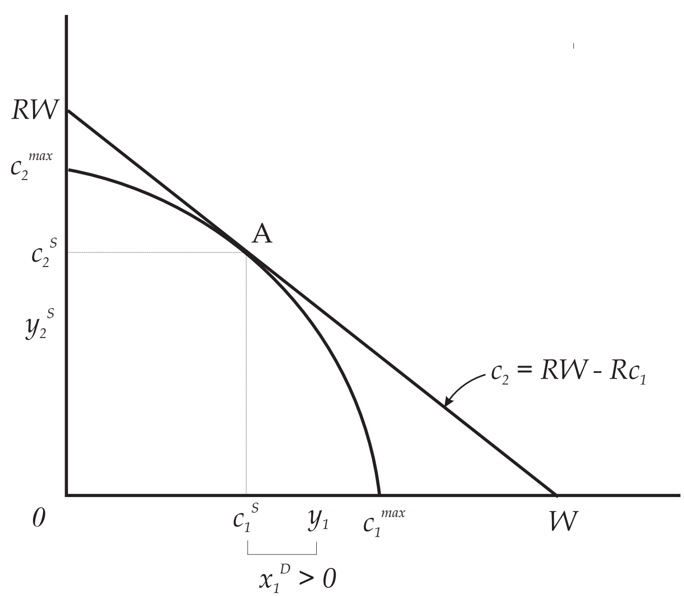
Observe that this equation is simply the PPF described in equation (6.4).

Crusoe's wealth (measured in units of current output) is the present value of his company's cash flow; i.e.,

$$W = c_1 + \frac{c_2}{R}. \tag{6.8}$$

Alternatively, this expression can be rewritten as $c_2 = RW - Rc_1$. As a business manager, Crusoe's problem boils down to choosing a sequence of 'cash flows' that maximizes the present value of the business subject to what is technologically feasible. In mathematical terms, he must choose (c_1, c_2) to maximize equation (6.8) subject to the constraint (6.7). The solution (c_1^S, c_2^S) is depicted as point A in Figure 6.5.

FIGURE 6.5
Stage 1: Maximizing Wealth



Since $x_1 = y_1 - c_1$, it follows that the investment demand function is given by $x_1^D = y_1 - c_1^S$. Notice that the slope of the PPF at point A is equal to $-R$. In other words, at point A it is true that $1 + MPK(k_1 + y_1 - c_1^S, z_2) = R$, which is equivalent to (6.6). This proves that the investment demand x_1^D characterized by condition (6.6) maximizes Crusoe's wealth.

- **Exercise 6.2.** Use Figure 6.5 to depict an allocation that lies on the PPF but does not maximize wealth.

- **Exercise 6.3.** Use Figure 6.5 to deduce how (c_1^S, c_2^S) and x_1^D respond to an exogenous increase in R. Explain. How does the increase in R affect wealth measured in present and future value? Explain.

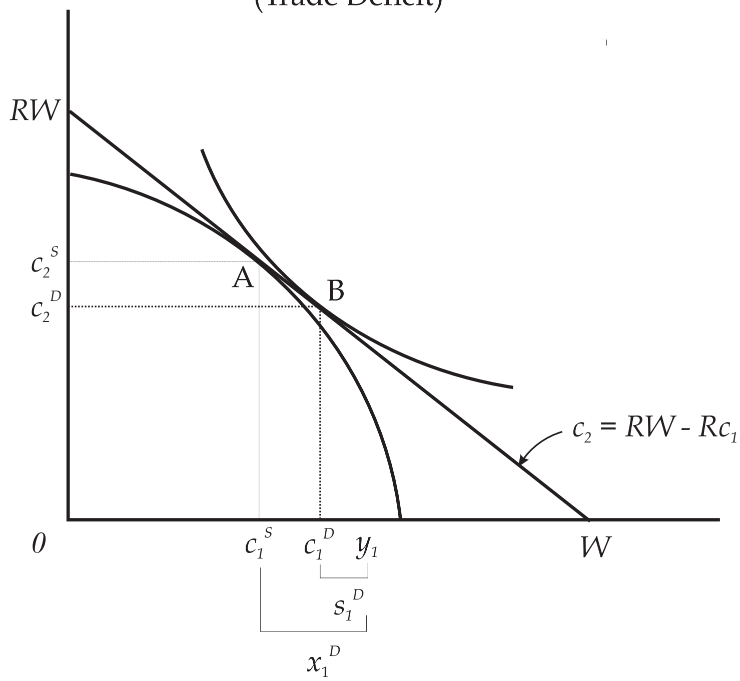
- **Exercise 6.4.** Reformulate the analysis above under the assumption that capital depreciates fully after it is used in production. Depict the solution in a diagram similar to Figure 6.5 and show that $MPK = R$ with $c_2^S = y_2^S = z_2 f(x_1^D)$.

The magnitude W derived above measures the value of Crusoe's business enterprise. In an open economy, his actual wealth may differ from W to the extent that he holds other financial assets (existing claims against foreigners) or liabilities (existing claims by foreigners against him). Let b_j denote Crusoe's net foreign bond holdings (financial claims against foreigners) as of the end of period $j = 0, 1, 2$. If b_j is positive, then Crusoe is a net creditor to the rest of world as of the end of period j. If b_j is negative, the Crusoe is a net debtor to the rest of world as of the end of period j. The magnitude b_0 denotes Crusoe's 'initial' foreign net asset position. In Chapter 4, we implicitly assumed that $b_0 = 0$. But if $b_0 \neq 0$, then Crusoe's wealth as of the beginning of period one is in fact given by $W + b_0$ and the first period GNP is given by $GNP = GDP + rb_0$.

6.4.2 Stage 2: Maximizing Utility

At this stage, Crusoe simply faces a standard choice problem that we have already studied at length in Chapter 4. That is, with wealth maximized at some level $W + b_0$, Crusoe's intertemporal consumption demands must maximize $u(c_1, c_2)$ while respecting the lifetime budget constraint: $c_1 + R^{-1}c_2 = W + b_0$. As in Chapter 4, we are free here to place the indifference curve any where along the intertemporal budget constraint (the exact position will depend on the nature of his preferences). The location of the indifference curve does not determine investment or GDP; what it determines is the actual consumption profile together with the current account position (net domestic saving position) of the economy. Figure 4.6 displays the solution (c_1^D, c_2^D) to Crusoe's utility maximization problem (under the assumption that $b_0 = 0$).

FIGURE 6.6
Stage 2: Maximizing Utility
(Trade Deficit)



Before proceeding, it will be useful to review some basic national income accounting for an open economy (without a government sector). First, recall the income-expenditure identity: $y_1 \equiv c_1 + x_1 + nx_1$, where nx_1 denotes the trade balance (net exports) in period one. The current account balance is defined as $ca_1 \equiv s_1 - x_1$ or $ca_1 \equiv (b_1 - b_0)$. The definition of national saving is given by $s_1 \equiv y_1 + rb_0 - c_1$. Combining these definitions, it follows that:

$$ca_1 \equiv nx_1 + rb_0.$$

In the special case where $b_0 = 0$, the current account surplus is precisely equal to the trade balance (in period one). As well, we have:

$$\begin{aligned} s_1 &\equiv x_1 + ca_1; \\ &\equiv x_1 + (b_1 - b_0); \end{aligned}$$

or $s_1 = x_1 + nx_1$ with $nx_1 = b_1$ if $b_0 = 0$.

Now, consider point B in Figure 6.6. Again, remember that the point B in Figure 6.6 could have been placed anywhere along the budget line. As drawn in Figure 6.6, Crusoe's economy has positive net savings in period one; i.e., $s_1^D = y_1 + rb_0 - c_1^D > 0$ (where recall that we have assumed $b_0 = 0$). These savings, however, are insufficient to finance the entire expenditure on domestic investment; i.e., $s_1^D < x_1^D$. The difference must be financed by importing output; i.e., $nx_1^D = s_1^D - x_1^D < 0$. These imports are paid for by issuing (selling) bonds to foreigners; i.e. $b_1^D = nx_1^D < 0$, so that in this example, Crusoe becomes a net debtor nation in period one. In Figure 6.6, the trade deficit corresponds to the distance $(c_1^D - c_1^S)$. This foreign debt is repaid (with interest) in period two, which is why $c_2^S > c_2^D$. Note that in period two, net exports are given by $nx_2^D = -(1+r)b_1^D = -(1+r)nx_1^D > 0$. The current account surplus in period two is given by $ca_2^D = nx_2^D + rb_1^D$. Since $b_1^D < 0$, it follows that $ca_2^D < nx_2^D$ (the future current account surplus is smaller than the future trade surplus).

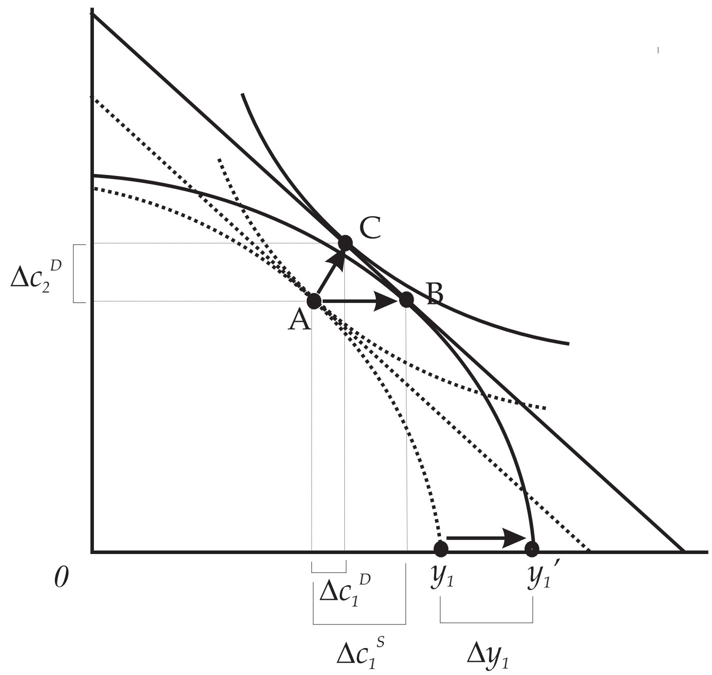
- **Exercise 6.5.** Using a diagram similar to Figure 6.6, depict an equilibrium for a small open economy in which the level of consumption is greater than current GDP. Depict another equilibrium in with the country runs a current account surplus.

6.4.3 A Transitory Productivity Shock

We are now ready to examine how our model economy reacts to a variety of shocks. Here, I will consider a transitory increase in productivity; i.e., $\Delta z_1 > 0$ and $\Delta z_2 = 0$. The experiment conducted here is similar to the one performed in Chapter 4 in the context of an endowment economy. To make drawing diagrams a bit easier, let me assume that capital depreciates fully after use and that $b_0 = 0$. None of the qualitative conclusions depend on these simplifications.

As before, let us begin with an initial situation in which the economy just happens (by coincidence) to be running a zero trade balance; this is depicted as point A in Figure 6.7. That is, initially we have $ca_1^D = nx_1^D = 0$. The effect of a transitory productivity shock is to shift the PPF in a parallel manner to the right; e.g., so that point A moves to point B. As the interest rate is exogenous, it remains unchanged, so that point B characterizes the business sector's optimal 'cash flow' (c_1^S, c_2^S). Assuming that both consumption goods are normal, their demands will increase along with the increase in wealth so that the new consumption allocation is given by point C.

FIGURE 6.7
A Transitory Productivity Shock



Recall that domestic investment demand is given by $x_1^D = y_1 - c_1^S$. Therefore, the change in domestic investment spending is given by: $\Delta x_1^D = \Delta y_1 - \Delta c_1^S$. From Figure 6.7, we see that $\Delta y_1 = \Delta c_1^S > 0$, so that $\Delta x_1^D = 0$. In other words, in a small open economy, a transitory productivity shock has no effect on domestic investment spending. This result is consistent with equation (6.6), which asserts that investment spending depends on the (expected) future productivity of capital z_2, but not on the current productivity of capital z_1. This makes sense since current investment becomes productive only in the future.

Now recall that desired national saving is given by $s_1^D = y_1 - c_1^D$. Therefore, the change in desired national saving is given by: $\Delta s_1^D = \Delta y_1 - \Delta c_1^D$. From Figure 6.7, we see that $\Delta y_1 > \Delta c_1^D > 0$, so that $\Delta s_1^D > 0$. This result is consistent with the consumption-smoothing argument that we discussed in Chapter 4.

Since current investment spending remains unchanged and since future productivity is the same, it follows that the future GDP remains unchanged; i.e., $\Delta y_2^S = \Delta c_2^S = 0$ (note that $y_2^S = c_2^S$ if capital depreciates fully after use). Since $\Delta c_2^D > 0$, future consumption now exceeds future GDP. This extra consumption is financed by $R\Delta s_1^D$; i.e., the principal and interest earned on the current increase in national saving.

Since $b_0 = 0$, we know that $ca_1 = nx_1$. We also know that $s_1^D = x_1^D + ca_1^D$, so that $\Delta s_1^D = \Delta x_1^D + \Delta ca_1^D$. Since we have already established that $\Delta x_1^D = 0$, it follows that $\Delta ca_1^D = \Delta s_1^D > 0$. In other words, the effect of this shock is to increase the current account surplus (and trade balance). That is, some of the period's extra GDP is used to augment current consumption with the rest used to purchase foreign bonds (which can later be cashed in to augment future consumption).

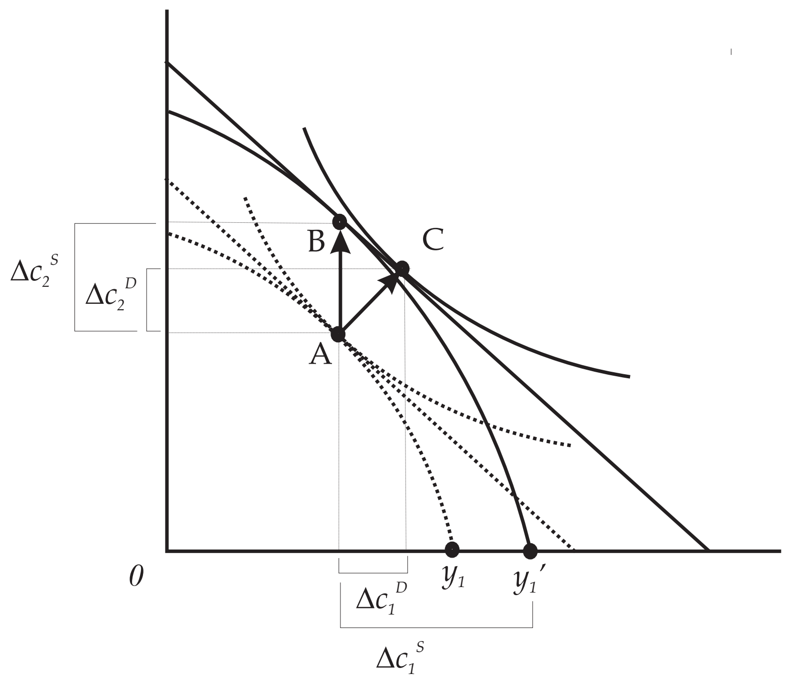
- **Exercise 6.6.** How would this small open economy respond to an anticipated increase in productivity? Explain, using a diagram similar to Figure 6.7.

6.4.4 A Persistent Productivity Shock

Let us now consider the effect of a persistent productivity shock; i.e., $\Delta z_1 > 0$ and $\Delta z_2 > 0$. Again, the experiment conducted here is similar to the one performed in Chapter 4 in the context of an endowment economy. Let us continue to assume that capital depreciates fully after use and that $b_0 = 0$.

As before, let us begin with an initial situation in which the economy just happens (by coincidence) to be running a zero trade balance; this is depicted as point A in Figure 6.8. That is, initially we have $ca_1^D = nx_1^D = 0$. A persistent productivity shock acts like a combination of a transitory and anticipated shocks. As such, the PPF shifts to the right (as in the previous experiment) and to the northwest (as in the previous exercise). Again, the real rate of interest remains unchanged, so that the business sector's optimal 'cash flow' (c_1^S, c_2^S) moves 'up' from point A to point B in Figure 6.8 (in fact, point B may end up either to the right or left of point A). As wealth is now higher, the pattern of consumer demands shifts to the northeast; i.e., from point A to point C.

FIGURE 6.8
A Persistent Productivity Shock



Recall that the change in domestic investment demand is given by: $\Delta x_1^D = \Delta y_1 - \Delta c_1^S$. From Figure 6.8, we see that $\Delta c_1^S > \Delta y_1 > 0$, so that $\Delta x_1^D > 0$. In other words, in a small open economy, a persistent productivity shock leads to an increase in domestic investment spending. This result is consistent with equation (6.6), which asserts that investment spending depends positively on the (expected) future productivity of capital z_2.

Now recall that the change in desired national saving is given by $\Delta s_1^D = \Delta y_1 - \Delta c_1^D$. From Figure 6.8, we see that $\Delta y_1 \approx \Delta c_1^D > 0$, so that the change in desired national saving is ambiguous (in Figure 6.8, I have drawn things such that $\Delta s_1^D \approx 0$). This result is consistent with the consumption-smoothing argument that we discussed in Chapter 4. That is, higher levels of income today induce people to save more. On the other hand, since individuals expect higher incomes in the future, they wish to save less today. Which effect dominates depends on the nature of preferences and the size of z_2 relative to z_1.

Since current investment spending increases along with (expected) future productivity, the model predicts that the (expected) future GDP should increase; i.e., $\Delta y_2^S = \Delta c_2^S > 0$. Notice here that $\Delta c_2^S > \Delta c_2^D > 0$, so that future consumption now falls short of future GDP. Evidently, some of the extra future output must be used to pay back the added foreign debt accumulated in period one.

Since $b_0 = 0$, we know that $ca_1 = nx_1$. We also know that $s_1^D = x_1^D + ca_1^D$, so that $\Delta s_1^D = \Delta x_1^D + \Delta ca_1^D$. Since we have already established that $\Delta x_1^D > 0$ and $\Delta s_1^D \approx 0$, it follows that $\Delta ca_1^D \approx -\Delta x_1^D < 0$. In other words, the effect of this shock is to reduce the current account surplus (and trade balance), so that the economy experiences a trade deficit. That is, since national savings do not change very much, the bulk of the extra investment spending is ultimately financed with imports of capital goods (the sale of bonds to foreigners).

6.4.5 Evidence

Mendoza (1991) documents the following facts for small open economies. First, the correlation between national saving and investment is positive. Second, national saving does not fluctuate as much as investment. Third, the trade balance is countercyclical (tends to move in the opposite direction of GDP over the cycle). He also shows that these basic facts are consistent with the predictions of a neoclassical model subject to persistent productivity shocks. While his model is considerably more complicated that the one developed above, the same basic idea there is at work here.

In particular, imagine that a persistent productivity shock hits our model economy. Imagine that the shock is such that $\Delta z_1 > \Delta z_2 > 0$ (i.e., the effect of the shock is strongest in the period it hits, but then tends to die out over time). According to our model, the anticipation of higher future productivity ($\Delta z_2 > 0$) should lead to an investment boom. The current productivity shock ($\Delta z_1 > 0$) results in an increase in current GDP (so that investment is procyclical, as it is in the data). Since the current shock is stronger than the anticipated shock, desired national saving should increase, but not by as much as the increase in desired investment. Thus, fluctuations in national saving should be smaller than fluctuations in investment. Furthermore, since saving does not increase as much as investment, the trade balance must fall (i.e., it moves in the opposite direction as GDP).

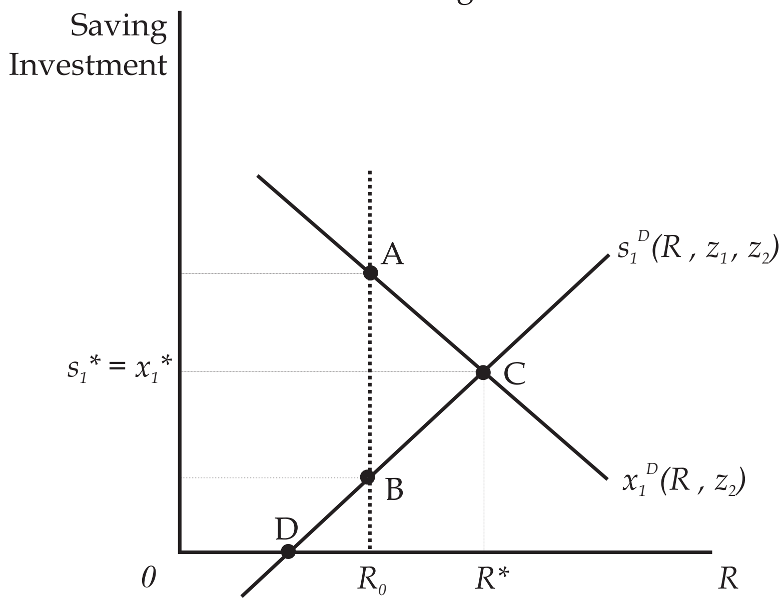
6.5 Determination of the Real Interest Rate

We have already discussed the determination of the interest rate in the context of the endowment economy studied in Chapter 4. Many of the basic insights developed there continue to hold here. Recall that in an open economy endowment economy, national saving is (by definition) given by:

$$s_1 \equiv ca_1.$$

In a closed economy, the interest rate must adjust to a point such that $ca_1^D(R^*) = 0$, so that $s_1^D(R^*) = 0$.

FIGURE 6.9
National Saving and Investment



In an open economy model with investment, however, national savings is (by definition) given by:

$$s_1 \equiv x_1 + ca_1.$$

In a closed economy, the interest rate must again adjust to a point such that $ca_1^D(R^*) = 0$. However, from the accounting definition above, we see that national savings can (and generally will be) positive, since savings can now be used to finance domestic investment expenditure. So, another way at looking at the determination of the interest rate is to think of R as adjusting to a point at which desired saving is just equal to desired investment; i.e.,

$$s_1^D(R^*, z_1, z_2) = x_1^D(R^*, z_2).$$

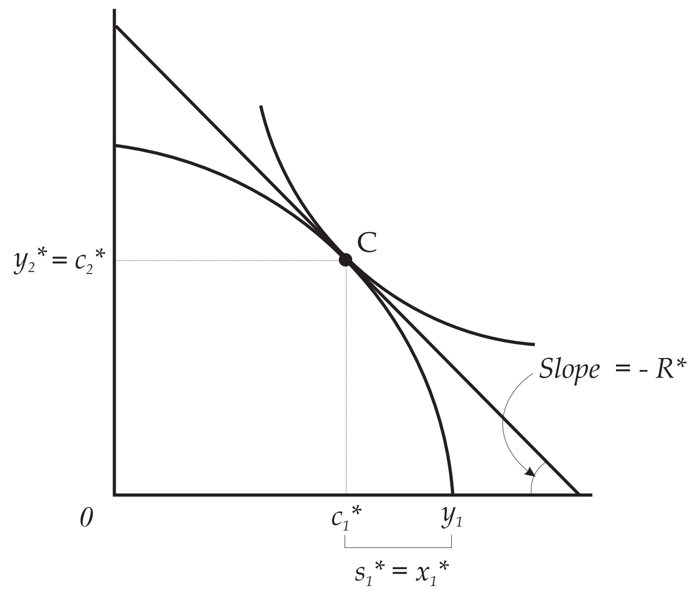
So, in terms of a diagram, we can think of depicting the economy as in Figure 6.9.

Suppose that the interest rate is exogenously determined by R_0 (small open economy scenario). Then the economy in Figure 6.9 will be running a current account deficit (the difference between points A and B). Note that point A corresponds to the same point A in Figure 6.6. Likewise, point B corresponds to the same point B in Figure 6.6.

Now, if this economy was closed to international trade, then there would be an excess demand for credit if the interest rate remained at R_0 (the demand for investment exceeds the supply of saving). The competition for loans would put upward pressure on the real interest rate. In a general equilibrium, the interest rate would be given by R^* (where the supply of saving just equals the demand for investment at point C).[1]

The general equilibrium can alternatively be characterized with a diagram similar to Figure 6.3 (except that here we are assuming that capital depreciates fully). In Figure 6.10, point C corresponds to point C in Figure 6.9.

FIGURE 6.10
General Equilibrium



Note that while the addition of capital and investment complicates the model beyond what was developed in Chapter 4, it does not alter the qualitative predictions concerning the effect of productivity shocks on the interest rate (show this as an exercise).

6.6 Summary

Capital is a durable asset which produces services that, together with labor, contributes to the production of an economy's GDP. The economy's capital

[1] Point D represents the general equilibrium in an endowment economy (where there is no investment demand function).

stock grows when the level of net investment is positive. The availability of an investment technology allows an economy to choose the intertemporal pattern of production. The availability of an investment technology also allows an economy to smooth its consumption patterns even in the absence of a financial market.

In a small open economy, the level of investment is determined primarily by the expected productivity of capital investment, together with the prevailing real interest rate. The current account position (and hence, net domestic saving) in a small open economy adjusts primarily to accommodate desired consumption patterns. As in the endowment economy, a 'deterioration' of a country's current account position may be associated with either an increase or decrease in the general level of welfare. Whether welfare improves or not depends on the nature of the shock hitting the economy. For example, a recession that leads to a transitory decline in GDP will lead to a decline in the current account position, as would an investment boom caused by a sudden improvement in the expected productivity of domestic capital spending.

To the extent that modern economies are integrated, the real rate of interest depends on both the world supply of credit and the world demand for investment. Shocks that affect large countries or large regions of the world may lead to a change in the structure of world interest rates. The model developed in this chapter should be viewed as presenting a rough guide as to the economic forces that are likely to influence the structure of real interest rates prevailing in world financial markets. One should keep in mind, however, that in reality there are many different types of interest rates. There are 'short' and 'long' rates; 'real' and 'nominal' rates; and 'risky and risk-free' rates. Some of these interest rates may be influenced primarily by local conditions.

6.7 Problems

1. Imagine that a small open economy is subject to transitory changes in productivity. Using a diagram similar to Figure 6.9, demonstrate how the supply of saving and the demand for investment functions fluctuate with these disturbances. Does the trade balance behave in a procyclical or countercyclical manner? Explain. Is this consistent with the evidence?

2. Repeat question 1, except now imagine that the productivity shocks are persistent.

3. If a shock hits the economy that sends the trade balance into deficit, does this necessarily imply that economic welfare declines? Explain.

4. Repeat questions 1 and 2 in the context of a closed economy and explain how the real interest rate is predicted to behave. Make sure to provide economic intuition for your results.

6.8 References

1. Mendoza, Enrique G. (1991). "Real Business Cycles in a Small Open Economy," *American Economic Review*, 81(4): 797–818.

Chapter 7

Labor Market Flows and Unemployment

7.1 Introduction

In this chapter, we take a closer look at the market. In previous chapters, we concerned ourselves with the determination of level of employment and cyclical fluctuations in the level of employment. Any change in the level of employment constituted *net* changes in employment (or hours worked). The data reveals, however, that net changes in employment are small relative to the gross flows of individuals that move into and out of employment. A zero net change in employment, for example, is consistent with one million workers finding jobs and one million workers losing jobs. In other words, individual employment patterns fluctuate a great deal more than the aggregate (the sum of all individual employment patterns). This chapter presents some data on gross worker flows and develops some simple theory to help us interpret this data.

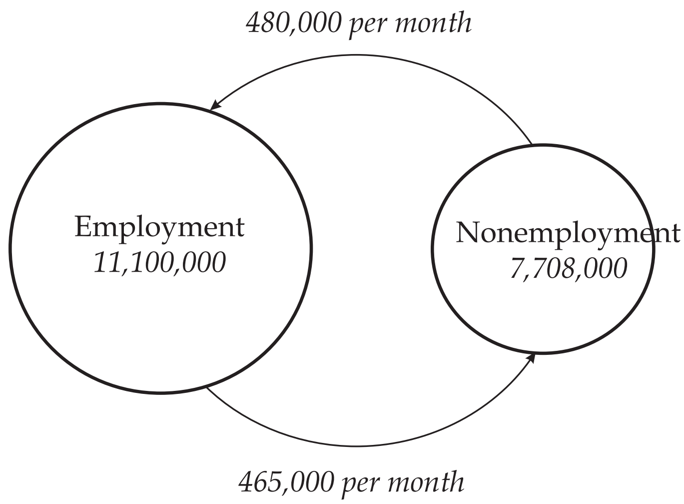
7.2 Transitions Into and Out of Employment

Many countries have statistical agencies that perform monthly labor force surveys that measure various aspects of labor market activity.[1] In these surveys, a person is labeled as *employed* if they report having done any work during the reference period of the survey. If a person is not employed, they are then labeled as *nonemployed*. Figure 7.1 summarizes the average stocks of employed and nonemployed individuals in Canada over the sample period 1976–1991 (see Jones, 1993). The figure also records the average *monthly* flows of workers into

[1] For example, the *Labour Force Survey* in Canada and the *Current Population Survey* in the United States.

and out of employment.

FIGURE 7.1
Average Labor Market Stocks and Flows
Canada 1976 - 1991 (Jones, 1993)



Over the sample period 1976–1991, the *net* monthly change in employment averaged only 15,000 persons (due mostly to population growth and an increase in female labor market participation). Notice how small the net change in employment is relative to the monthly *gross* flows; i.e., in a typical month, *almost one million* individuals flow into or out of employment. The existence of large gross flows that roughly cancel each other out is evidence that individuals are subject to *idiosyncratic shocks* (changes in individual circumstances) that roughly cancel out in the aggregate. In other words, even in the absence of any *aggregate shocks* (a shock that effects most people in the same way–as in earlier chapters), it appears that individuals are subject to a considerable amount of uncertainty in the labor market.

How are we to interpret the apparent 'instability' of employment (from an individual's perspective)? Is the labor market simply a huge game of 'musical chairs?' Should the government undertake job creation programs (i.e., attempt to increase the number of 'chairs') to reduce the amount of labor market turnover? Is labor market turnover a good thing or a bad thing? To answer these questions, we need to develop a model of labor market transitions.

7.2.1 A Model of Employment Transitions

Consider a model economy consisting of a fixed number of individuals who have preferences defined over consumption and leisure (c, l) given by:

$$u(c, l) = \ln(c) + vl, \tag{7.1}$$

where $v \geq 0$ can be interpreted as either a preference parameter measuring the value of leisure, or as a productivity parameter measuring the productivity of a home-production technology. Assume that people differ in their 'leisure value' parameter v. Notice that the utility of consumption c is given by the natural logarithm $\ln(c)$. The natural logarithm is an increasing and concave function with the property that as c approaches zero, $\ln(c)$ approaches negative infinity. These properties imply the following reasonable features associated with the utility of consumption: (1) utility is increasing in the level of consumption (more is preferred to less); (2) utility increases at a decreasing rate (diminishing marginal utility of consumption); and (3) very low levels of consumption are very painful (utility approaches negative infinity).

People also differ in their skill levels. If the market price of one's labor is related to one's skill (a reasonable assumption), then individuals will also face different *potential* wage rates w. Let me also assume (again, quite reasonably) that individuals have different levels of financial wealth, from which they generate nonlabor income a. Thus, at any point in time, an individual is characterized by an endowment (w, a, v). Assume that individuals face *idiosyncratic risk* in their endowments, so that elements of (w, a, v) fluctuate randomly over time. Finally, assume that these risks cancel out in the aggregate (so that there is no aggregate risk).

As in Chapter 2, individuals are endowed with one unit of time. Let n denote the time that an individual allocates to the labor market. Then individuals are assumed to face the budget constraint:[2]

$$c = wn + a. \tag{7.2}$$

Inserting this budget constraint into (7.1) together with the time constraint $l = 1 - n$ allows us to rewrite the objective function as:

$$V(n) = \ln(wn + a) + v(1 - n). \tag{7.3}$$

Hence, the individual's choice problem boils down to choosing an appropriate allocation of time n (just as in Chapter 2).

For simplicity, assume that time is indivisible, so that individuals must specialize in their use of time; i.e., $n = 0$ (nonemployment) or $n = 1$ (employment). Then the utility payoff from employment is given by $V(1) = \ln(w + a)$ and the

[2] Implicit in this budget constraint is the assumption that individuals cannot save or borrow and that there are no insurance markets.

utility payoff from nonemployment is $V(0) = \ln(a) + v$. The optimal employment decision is therefore given by:

$$n^* = \begin{cases} 1 & \text{if} \quad V(1) \geq V(0); \\ 0 & \text{if} \quad V(1) < V(0). \end{cases}$$

Figure 7.2 plots the $V(1)$ and $V(0)$ as a function of the wage w (for a given a and v).

FIGURE 7.2
Work versus Leisure

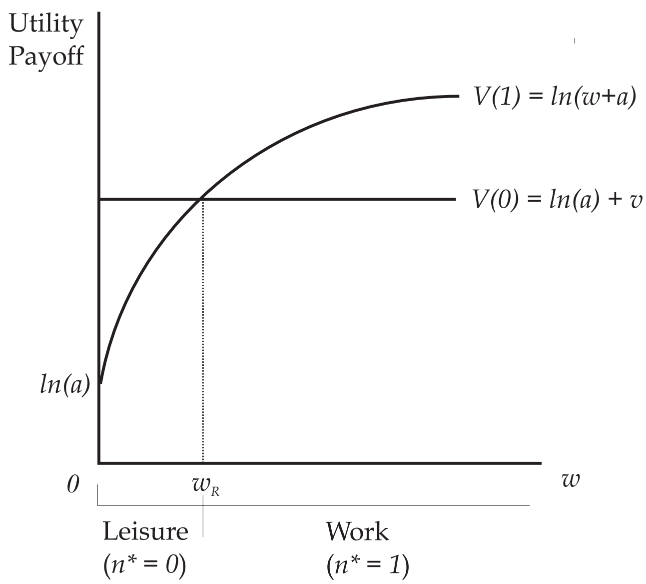


Figure 7.2 reveals a common sense result: individuals whose skills command a high price in the labor market are more likely to be working (holding all other factors the same). In particular, individuals with $w < w_R$ will choose *leisure*, while those with $w \geq w_R$ will choose *work*. (Technically, those with $w = w_R$ are just indifferent between working or not, but we can assume that when indifferent, individuals choose work).

The wage w_R is called the 'reservation wage.' The reservation wage is determined by the intersection of the functions $V(1)$ and $V(0)$ in Figure 7.2, so that w_R solves the equation:

$$\ln(w_R + a) = \ln(a) + v. \tag{7.4}$$

We can solve[3] equation (7.4) for w_R; i.e.,

$$w_R = (e^v - 1)a. \tag{7.5}$$

[3] Recall the following properties of logarithms: $\ln(e^x) = x\ln(e) = x$ (since $\ln e = 1$); and $\ln(xy) = \ln(x) + \ln(y)$.

Notice that the reservation wage is a *function* $w_R(a, v)$; i.e., it depends positively on both a and v.

The reservation wage has a very important economic interpretation. In particular, it represents the price of labor for which an individual is just indifferent between working or not. In other words, it is the minimum wage that would induce an individual to work. As such, the reservation wage is a measure of an individual's 'choosiness' over different wage rates. That is, an individual with a high reservation wage is someone who is very choosy, while someone with a low reservation wage is not very choosy. What determines an individual's degree of choosiness over job opportunities? The reservation wage function in (7.5) tells us that there are two primary factors that determine choosiness: (1) the level of non-labor income (a); and (2) the value of time in alternative uses (v). Choosy individuals are those with either high levels of wealth or those who attach great value to non-market activities.

Notice that the individual's labor supply function can also be expressed in terms of their reservation wage; i.e.,

$$n^* = \begin{cases} 1 & \text{if} \quad w \geq (e^v - 1)a; \\ 0 & \text{if} \quad w < (e^v - 1)a. \end{cases}$$

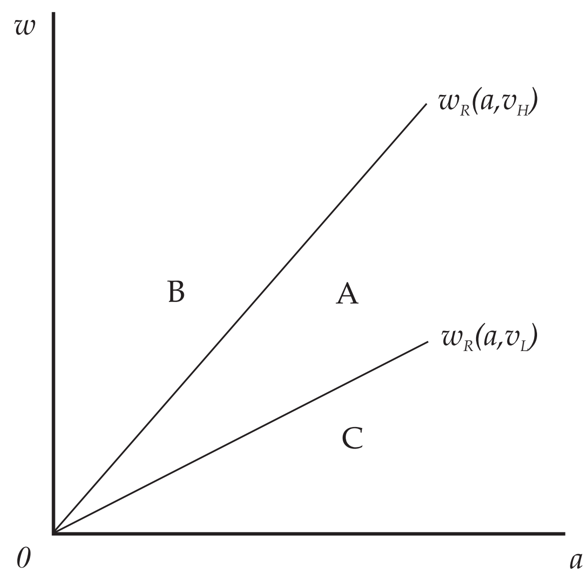
Expressing the labor supply function in this way makes it clear that labor supply tends to be increasing in w, but decreasing in both a and v (higher levels of a and v make people more choosy and therefore less likely to work at any given wage).

Our theory also tells us how each person's economic welfare (maximum utility level) depends on their endowment (w, a, v). In particular, an individual's welfare is given by $W(w, a, v) = \max\{\ln(w + a), \ln(a) + v\}$; i.e., the maximum of either $V(1)$ and $V(0)$. In Figure 7.2, W is just the 'upper envelope' of the functions $V(1)$ and $V(0)$. According to our theory, the welfare function is (weakly) increasing in w, a and v. What this means is that it is impossible for an increase in any of these parameters to make an individual worse off (and will, in general, make them better off). An important implication of this result is that there is no straightforward way of linking a person's employment status with their level of welfare. Likewise, we cannot generally make statements about how two economies are performing relative to each other simply by looking at employment levels.

- **Exercise 7.1.** Consider two economies A and B that are identical in every respect except that the individuals in economy A have greater levels of wealth (as measured by the parameter a). According to our theory, which economy is likely to exhibit the higher level of employment? In which economy are individuals likely to be better off? Explain.

Figure 7.3 plots the reservation wage function $w_R(a, v)$ in (w, a) space for two types of individuals who differ in v ($v_H > v_L$). From equation (7.5), note that the slope of the reservation wage function is given by $(e^v - 1)a \geq 0$.

FIGURE 7.3
Reservation Wage Functions



An individual with a particular endowment (w, a) is located at some point in the space depicted in Figure 7.3. Imagine that individuals are located at various points in this space. Then type v_H individuals located in region B will be employed, while those located in regions A and C will be nonemployed. Likewise, type v_L individuals located in regions A and B will be employed, while those located in region C will be nonemployed.

- **Exercise 7.2.** Consider two individuals (Bob and Zu) who are located precisely at the point A in Figure 7.3 (i.e., they both have identical labor market opportunities and identical wealth levels). Bob has $v = v_L$ while Zu has $v = v_H$. Which of these two people will be employed and which will have a higher level of utility? Explain.

According to our theory, the aggregate level of employment is determined by how individuals are distributed across the space (w, a, v). To the extent that this distribution remains constant over time, so will the aggregate level of employment (there will be no net changes in employment over time). However, to the extent that individuals experience changes in (w, a, v), the economy will, in general, feature gross flows of individuals into and out of employment similar to what is observed in Figure 7.1. For example, consider an individual with $v = v_L$ who is initially located at point A in Figure 7.3. Suppose that this individual experiences a decline in the demand for their skill, so that the market price of their labor falls (the individual moves to point C). Then this individual will (optimally) make a transition from employment to nonemployment (and the person will be made worse off in this case, owing to the exogenous decline in the market value of his labor).

- **Exercise 7.3.** Consider two economies that are identical in every way except that in one economy all individuals have zero wealth ($a = 0$). Which economy will the higher level of output and employment? Which economy will feature a higher degree of labor market turnover? Which economy would you rather live in? Explain.

7.3 Unemployment

None of the models we have studied so far can explain the phenomenon of unemployment. Some people have the mistaken impression that an unemployed person is someone who 'wants' to work at 'the' prevailing wage rate, but for some reason cannot find a job. Such a definition is problematic for a number of reasons. First, since individuals obviously differ in skills (among other traits), it is very difficult to identify what 'the' prevailing wage might be for any given individual. An individual may claim to be worth \$20 an hour but may in fact be worth only \$10 an hour. What would it mean for such an individual to claim that they cannot find a job (that pays \$20 an hour)?

But more importantly, this is not the way unemployment is defined in labor market surveys. A labor market survey first asks a person whether they are working or not. If they are working (or have worked in the reference period of the survey), they are labeled as employed. If they report that they are not working, the survey then asks them what they did with their time by checking the following boxes (item 57 in the Canadian *Labor Force Survey*):

- **57** IN THE PAST 4 WEEKS, WHAT HAS ... DONE TO FIND WORK? (Mark all methods reported):

  ☐ NOTHING;
  ☐ PUBLIC employment AGENCY;
  ☐ PRIVATE employment AGENCY;
  ☐ UNION;
  ☐ FRIENDS or relatives;
  ☐ Placed or answered ADS;
  ☐ LOOKED at job ADS;
  ☐ OTHER, Specify in NOTES.

The *Current Population Survey* in the United States asks a similar set of questions. In Canada, if a nonemployed person checks off 'nothing,' then they are labeled *nonparticipants* (or not in the labor force). In the United States, if a person checks either 'nothing' or 'looked at job ads,' they are labeled as nonparticipants. If any other box is checked, then the person is labeled as *unemployed*. Clearly, a person is considered to be unemployed if: (1) they are nonemployed; and (2) if they are 'actively' searching for employment. In Canada, simply 'looking at job ads' is considered to be 'active' job search, while in the United States it is not.

Notice that the survey never actually asks anyone whether they are unemployed or not. Similarly, the survey does not ask whether people 'want' to work but were unable to find work. For that matter, the survey also does not ask people whether they 'want' leisure but were unable to find leisure (arguably a much more relevant problem for most people). Thus, among the group of nonemployed persons, the unemployed are distinguished from nonparticipants on the basis of some notion of active job search. Figure 7.4 provides some data for Canada over the sample period 1976–1991 (again, from Jones, 1993).

FIGURE 7.4
Average Labor Market Stocks and Monthly Flows
Canada 1976 - 1991 (Jones, 1993)

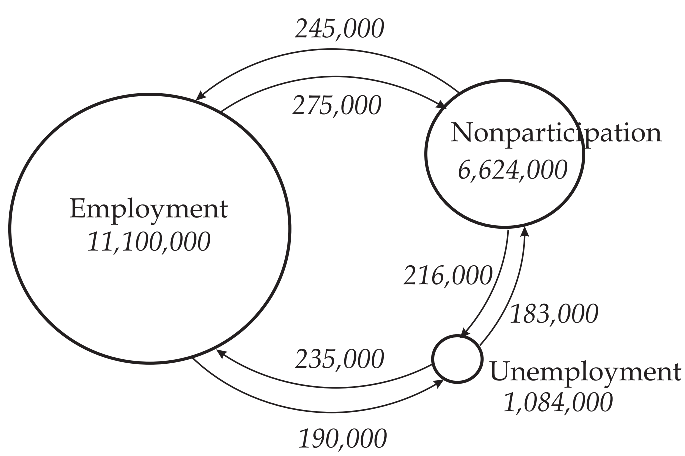


Figure 7.4 reveals a number of interesting facts. First, observe that over half of all individuals who exit employment in any given month become nonparticipants, rather than unemployed (i.e., the exit the labor force, which is defined to be the sum of employment and unemployment). Second, note that over half of all individuals who find employment in any given month were not unemployed

(i.e., they find work as nonparticipants). This latter fact casts some doubt on the empirical relevance of the concept of unemployment (nonemployed persons who actively search for work). On the other hand, note that the monthly probability of becoming employed is much greater for the unemployed $(235/1084) = 0.217$ than for nonparticipants $(245/6624) = 0.037$. This fact lends support to the notion that the unemployed are more intensively engaged in job search activities.

The models that we have studied to this point are ill-equipped to deal with the issue of unemployment (at least, as the concept is defined by labor force surveys). The reason for this is because there is no need for our model individuals to engage in job search activities. In those models, including the one developed in the previous section, everyone knows where to get the best value (highest wage) for their labor. They may not be happy about the going wage for their labor, but given this wage the choice is simply whether to allocate time in the labor market or allocate time to some other activity (like home production or leisure). In order to explain unemployment, we have to model a reason for why people would willingly choose to allocate time to an activity like job search.

7.3.1 A Model of Unemployment

Why do people search for jobs? For that matter, why do people search for anything (like mates)? It seems apparent that the key friction that generates search behavior is *imperfect information* concerning the location of one's ideal job match (or mate). In the model developed above, everyone knows where the location of their best job presides; this job pays w. The time allocation choice therefore only entails the question of whether or not to exploit the job opportunity that pays w (or equivalently, whether to exploit the home opportunity that 'pays' v).

In order to model job search, let me reformulate the model developed above in the following ways. First, let us assume (in order to simplify matters) that preferences are given by:

$$u(c,l) = c + vl.$$

Individuals have one unit of time that they can still allocate either to work or leisure. But now, there is a third option for their time: job search. Job search occurs at the beginning of the period. Assume that the job search activity consumes a fraction $0 < \theta < 1$ of the available time endowment. At the end of the job search process, an individual must then choose to allocate any remaining time $(1-\theta)$ to either work or leisure.

Let us assume then that individuals begin each period with a job opportunity that pays w, but that generally speaking, individuals realize that better job opportunities exist 'out there somewhere.' Locating these better job opportunities takes time and the outcome of the search process is uncertain (individuals may or may not come across better job opportunities when they search). Of course, feel free to substitute the word 'job' with 'mate;' in which case we have an explanation for why some people choose to be single (at least temporarily) and search for mates.

For simplicity, assume that in order to search, an individual must forever abandon his/her current job opportunity w (there is no going back to your old boyfriend if you dump him). Let w' denote the wage that will be paid at the new job opportunity. During the search process, this new job is 'located' with probability $(1-\pi)$, where $0 < \pi < 1$ denotes the probability of failure. Remember that in the event of failure, individuals are free to allocate the remainder of their time θ to leisure.

We must now try to discern which people choose to work, search or leisure. We proceed in the following way. Suppose, for the moment, that search is not an option. Then the reservation wage for individuals is given by $w_R = v$ (i.e., this is the wage that equates the utility payoff of work and leisure: $w_R + a = a + v$). Label those individuals who prefer work to leisure as *type A* individuals. Likewise, label those individuals who prefer leisure to work as *type B* individuals.

We know that type A individuals prefer work to leisure. But if they are now presented with a search option, which of these individuals prefer work to search? The payoff to work is given by w. The (expected) payoff to search is given by $(1-\theta)[(1-\pi)w' + \pi v]$. We can identify a reservation wage w_R^S that identifies the type A individual who is just indifferent between work and search; this reservation wage is given by:

$$w_R^S = (1-\theta)[(1-\pi)w' + \pi v]. \tag{7.6}$$

Thus, any type A individual with $w < w_R^S$ will prefer search over work. This reservation wage is depicted in Figure 7.5 in (w, v) space.

Let us now consider type B individuals. We know that these individuals prefer leisure to work. But now that they have a search option, which of these individuals will prefer search to leisure? The payoff to leisure is given by v. The (expected) payoff to search is given by w_R^S. Consequently, we can identify a reservation 'leisure value' v_R that just equates the payoff to leisure and search; i.e.,

$$v_R = (1-\theta)[(1-\pi)w' + \pi v_R].$$

We can solve the expression above for v_R as follows:

$$v_R = \left[\frac{(1-\theta)(1-\pi)}{1-(1-\theta)\pi}\right] w', \tag{7.7}$$

and is plotted in Figure 7.5. Here, any type B individual with $v < v_R$ will prefer search to leisure.

FIGURE 7.5
Work, Search and Leisure

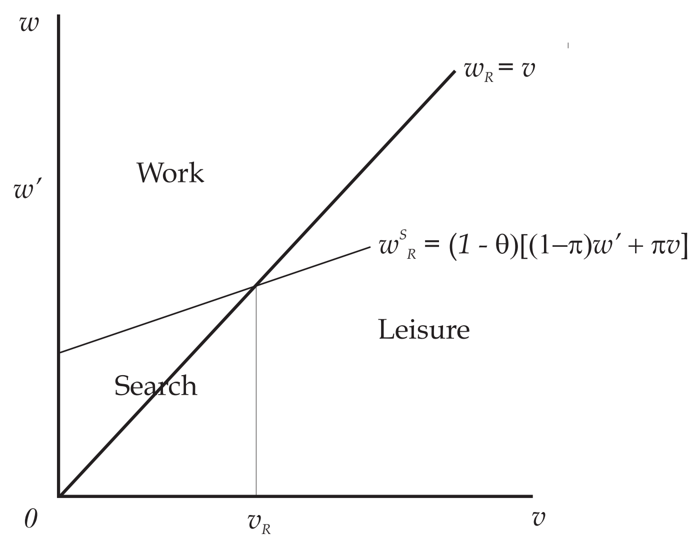


Figure 7.5 reveals the following. The individuals who are most likely to search are those who are presently poorly endowed in terms of both their present job opportunity *and* home opportunity (i.e., low values of w and v). For these individuals, allocating time to search is not very expensive (in terms of opportunity cost). The individuals who are most likely to work are those who currently have a good job opportunity (w) *and* a comparative advantage in working (w is high relative to v). The individuals who are most likely to choose leisure (nonparticipants) are those who have a good home opportunity (v) *and* a comparative advantage in leisure (v is high relative to w).

The model developed above here capable of generating labor market flows between employment, unemployment, and nonparticipation. These flows are triggered by changes in individual circumstances (i.e., changes in w and v). At any point in time, some individuals have sufficiently poor market and nonmarket opportunities that allocating time to search activity makes sense. Note that not everyone who searches would be picked up by the *Labour Force Survey* as being unemployed. In the model, the individuals who would be classified as unemployed are those who search *and* are unsuccessful.

Finally, we know that individual welfare is generally increasing in both w and v. Since those people who choose to search are those with very low w and v, we can conclude that the unemployed are generally among the least well-off in society. However, it is important to keep in mind that these people are less well-off *not* because they are unemployed but because they are endowed with low w and v. In particular, the concept of 'involuntary unemployment' makes no sense (since people obviously have a choice whether to search or not). On the other hand, it may make sense to think of some people as being involuntarily endowed with poor skills or poor opportunities in the home sector. Since the choice to search is voluntary, it follows that some level of unemployment (single people) is optimal. For example, a government could in principle eliminate unemployment by forcing people to work (a policy adopted in some totalitarian regimes). While measured unemployment would fall to zero, one would be hard pressed to argue that economic welfare must therefore be higher.

- **Exercise 7.4.** Using Figure 7.5, locate two individuals A and B such that A chooses to work while B chooses to search, but where B is better off than A (in terms of expected utility). Explain.

7.3.2 Government Policy

If individuals are risk-averse and if they are exposed to uncertainty in how their economic circumstances evolve over time (i.e., random changes in w, a and v), then they will want to insure themselves against such risk. Unfortunately, markets that would allow individuals to insure themselves against changes in the value of their human capital (i.e., changes in either w or v), either do not exist or function poorly. The reason for why this is so is a matter of debate. One view holds that because the true value of human capital is known only to the individual (i.e., it is *private information*), individuals may have the incentive to lie about the true value of their human capital just to collect insurance. For example, if a person becomes unemployed, it is not clear whether he is unemployed for 'legitimate' reasons (i.e., a true drop in w), or whether he is simply choosing not to work and simply *reporting* a drop w. Since w is private information, the insurance company has no way of knowing the truth and hence no way of providing an insurance policy that pays for itself.[4]

In such an environment, there may be a role for government provided insurance. Although the government must presumably cope with the same information frictions that afflict private insurance markets, the government does have one advantage over private firms. In particular, the government has the power of coercion so that it can make participation mandatory and collect 'fees' by way of taxes.[5] By operating a well-designed insurance scheme, the government may be able to improve economic welfare; see Andolfatto and Gomme (1996).

[4] Another view holds that the source of the friction is a government policy that guarantees the inalienability of human capital, thereby preventing insurance companies from garnisheeing the returns to human capital; i.e., see Andolfatto (2002).

[5] While governments make it difficult for private firms to garnishee human capital, the government itself does not restrict itself in this manner; see the previous footnote.

7.4 Summary

Aggregate fluctuations in the aggregate per capita labor input are generally not very large. However, this apparent stability at the aggregate level masks a considerable degree of volatility that occurs at the individual level. Modern labor markets are characterized by large gross flows of workers into and out of employment, as well as large gross flows into and out of other labor market states, like unemployment and nonparticipation. These large gross flows suggest that if policy is to be desired at all, it should likely be formulated in terms of redistributive policies (like unemployment insurance and welfare), rather than aggregate 'stabilization' policies (unless one takes the view that business cycles are caused by 'animal spirits').

The notion of 'unemployment' is a relatively modern concept, evidently emerging sometime during the Industrial Revolution (c. 1800). Unemployment is sometimes viewed as the existence 'jobless' workers or 'individuals who want work.' Unemployment rate statistics, on the other hand, define the unemployed at those individuals who are not working by actively searching for work. The distinction is important because nonparticipants are also technically 'jobless.' And the concept of 'wanting work' does not make sense since work is not a scarce commodity. What is scarce are relevant skills (which largely determine the market price of one's labor). To the extent that active job search constitutes a productive investment activity, the notion that measured unemployment represents 'wasted' or 'idle' resources (as is sometimes claimed) may not be appropriate.

Judgements about the economic welfare of individuals or economies made on the basis of labor market statistics like employment or unemployment must be made with care. Economic well-being is better measured by the level of broad-based consumption. The level of consumption attainable by individuals depends on a number of individual characteristics, including skill, age, heath, work ethic and wealth. The overall level of productivity (technology) and government policies (taxes, trade restrictions, etc.), also have a direct bearing on individual well-being. The choices that individuals make in the labor market are driven primarily by their individual characteristics. Changes in these characteristics may trigger labor market responses that do not vary in any systematic way with economic welfare.

The individual characteristics that lead individuals to be unemployed are typically such that the unemployed constitute some of the least fortunate members of a society. However, one should keep in mind that most societies are also made up of individuals who may be labelled the 'working poor.' Many nonparticipants are also not particularly well off. By narrowly focussing policies to help the unemployed, one would be ignoring the plight of an even larger number of individuals in need. To the extent that private insurance markets do not work perfectly well, there may be a role for a government 'consumption insurance' policy to help those in need (be they unemployed, employed or nonparticipants).

7.5 Problems

1. Consider the utility function (7.3). The MRS for these preferences is given by $MRS = vc$. Suppose that individuals are free to choose any $0 \leq n \leq 1$. Assume that $v \geq 1$. Show that an individual's employment choice is given by $n^* = \max\left\{\frac{w-va}{vw}, 0\right\}$. Find an expression for this person's reservation wage. Consider two individuals that differ in a (e.g., $a_H > a_L$). Assume that $va_L < w < va_H$. Depict the choices that these people make using an indifference curve and budget line drawn in (c, l) space.

2. Many economies have instituted minimum wage laws; i.e., laws that prevent individuals from working at jobs that pay less than some mandated wage w_M. Using the theory associated with Figures 7.2 and 7.3, explain the economic and welfare consequences of such a law. Does a minimum wage lead to unemployment in this model? Explain.

3. Recent technological advances in information technology (IT), for example, *Ebay* and *Workopolis.com*, have led to an increase in the efficiency of matching technologies. In the context of the model developed in this chapter, one might capture such technological advances by supposing that $p_A(v) > p_B(v)$ and $q_A(v) > q_B(v)$. According to our theory, how is such a technological advancement likely to affect the equilibrium level of vacancies and the natural rate of unemployment?

4. Figure 7.7 (see Appendix 7.A) was drawn under the assumption that $0 < (1-s-p) < 1$. However, it is also possible that $-1 < (1-s-p) < 0$. Redraw Figure 7.7 under this latter assumption and show how, for any given initial level of unemployment, the equilibrium dynamic path of unemployment 'oscillates' as it approaches the natural rate of unemployment.

7.6 References

1. Andolfatto, David (2002). "A Theory of Inalienable Property Rights," *Journal of Political Economy*, 110(2): 382-393.

2. Andolfatto, David and Paul A. Gomme (1996). "Unemployment Insurance and Labor Market Activity in Canada, *Carnegie-Rochester Conference Series on Public Policy*, 44: 47–82.

3. Jones, Stephen R. (1993). "Cyclical and Seasonal Properties of Canadian Gross Flows of Labour," *Canadian Public Policy*, 19(1): 1–17.

4. Pissarides, Christopher A. (1985). "Short-Run Equilibrium Dynamics of Unemployment, Vacancies and Real Wages, *American Economic Review*, 75(4): 676–690.

7.A A Dynamic Model of Unemployment

In this appendix, I describe a simplified version of a popular model introduced by Pissarides (1985). Imagine a world with a fixed number of individuals who either work if they have a job or search if they do not (i.e., assume that $v = 0$). The economy is also populated by firms that either produce output (if they have a worker) or recruit workers (i.e., are vacant) if they do not. Each firm requires only one worker and a firm-worker pair produce a level of output equal to y.

Firm-worker pairs negotiate a wage payment that divides the output y between them. Let $\pi = \theta y$ denote the profit accruing to the firm (so that $(1-\theta)y$ is the wage paid to the worker), where $0 < \theta < 1$ is now a parameter that indexes the bargaining power of the firm. After producing output in the current period, the firm-worker match survives into the next period with probability $(1 - s)$, where $0 < s < 1$ is an exogenous probability of separation (the probability that some shock occurs that results in the firm shutting down). If the firm-worker pair survive into the next period, they produce output (and split it) as before. If the firm-worker match breaks down, then the firm becomes vacant and the worker becomes unemployed. If we let r denote the real (net) rate of interest, then the capital value (i.e., the present value of the expected stream of profits) is given by:

$$J = \frac{\theta y}{r + s}. \tag{7.8}$$

- **Exercise 7.5.** Explain (i.e., do not simply describe) how the value of a firm depends on the parameters θ, y and s.

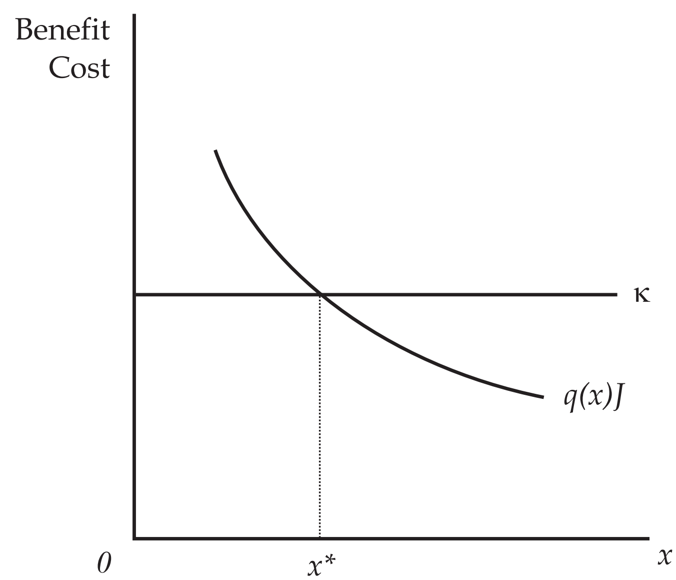
Since there is no centralized labor market, vacant firms and unemployed workers must seek each other out in a 'matching market.' A vacant firm must pay the cost κ to enter this market, but unemployed workers are let in for free (feel free to interpret vacant firms as 'single men,' unemployed workers as 'single women,' and the matching market as a 'nightclub'). Once inside the 'nightclub,' the matching technology works as follows. Let x denote the number of vacant firms (that choose to pay the entrance fee). Then a vacant firm matches with an unemployed worker with probability $q(x)$. Assume that q is a decreasing function of x, which implies that a greater number of vacancies increases competition among searching firms and so reduces the chances of any given vacancy from 'making contact' with an unemployed worker. An unemployed worker, on the other hand, finds a suitable vacancy with probability $p(x)$, where p is an increasing function of x (the greater the number of men, the better are the chances for the ladies).

If a vacant firm meets an unemployed worker, then they begin to produce beautiful output together. Thus, from the perspective of the vacant firm, the expected gain from paying the recruiting cost κ is given by $q(x)J$. If $q(x)J > \kappa$, then it would be worthwhile for more unmatched firms to incur the cost κ, which would then lead to an increase in the number of vacancies x. But as x increases, the probability of a successful match falls. Imagine that x increases to the point x^* such that an unmatched firm is just indifferent between paying the entrance cost κ and not; i.e.,

$$q(x^*)J = \kappa. \tag{7.9}$$

The condition above is depicted graphically in Figure 7.6.

FIGURE 7.6
Determination of Vacancies



- **Exercise 7.6.** Using conditions (7.8) and (7.9), show that an exogenous increase in productivity (y) leads to an increase in recruiting intensity (vacancies). Depict the change in a diagram similar to Figure 7.6. Explain the economic intuition behind this result.

The final thing to show is what this theory implies for the evolution of the level of unemployment over time. The level of unemployment at any given point in time t is given by u_t. If we let L denote the labor force, then the level of employment is given by $L - u_t$. Since L is a constant, we are free to set L to any number; e.g., $L = 1$ (so that u_t now represents the unemployment rate). Over time, the unemployment rate must evolve according to:

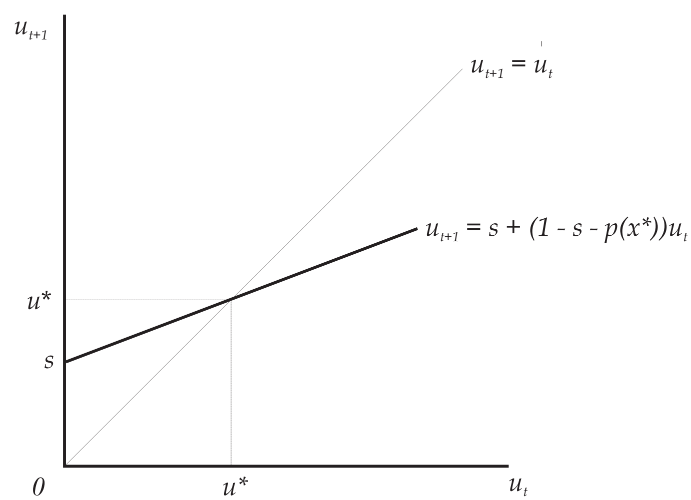
Future Unemployment = Current Unemployment + Job Destruction - Job Creation;

or,

$$\begin{aligned} u_{t+1} &= u_t + s(1 - u_t) - p(x^*)u_t; \\ &= s + (1 - s - p(x^*))u_t. \end{aligned} \tag{7.10}$$

Equation (7.10) is depicted in Figure 7.7 (assuming that $0 < p + s < 1$).

FIGURE 7.7
Equilibrium Unemployment
Rate Dynamics and Steady State



- **Exercise 7.7.** Use Figure 7.5 to show that for any given initial unemployment rate u_0, that the equilibrium unemployment rate converges to a steady-state unemployment rate u^*.

The unemployment rate u^* in Figure 7.7 is called a 'steady-state' unemployment rate because once this point is reached, the economy will stay there forever (assuming that nothing else changes). Sometimes, u^* is called the *natural rate of unemployment* (NRU). We can use equation (7.10) to solve for the NRU; i.e., setting $u_{t+1} = u_t = u^*$, we can calculate:

$$u^* = \frac{s}{s + p(x^*)}. \tag{7.11}$$

- **Exercise 7.8.** In many countries, the government levies a payroll tax on firms and uses the proceeds to pay unemployed workers unemployment benefits. Using the theory developed above, explain what effect such a payroll tax is likely to have on the natural rate of unemployment (hint: a payroll tax will reduce J).

- **Exercise 7.9.** Data from many countries reveals that there is a negative relationship between unemployment and vacancies; see Figure 7.8. This negative relationship is referred to as the 'Beveridge Curve.' Is the theory developed above consistent with the observed Beveridge curve relationship in the data?

Figure 7.8
Canadian Unemployment and
Vacancy Rates
1966 - 1988

Part II

Macroeconomic Theory: Money

Chapter 8

Money, Interest, and Prices

8.1 Introduction

None of the theory developed to this point has been concerned with explaining the behavior of *nominal* variables, like the nominal GDP, the nominal wage, and the price-level. This is because the theory developed above abstracts from any concept of money. Since nominal variables constitute values measured in units of money, the absence of money in any model necessarily prevents one from addressing nominal phenomena.

To many people, it seems surprising to think that one can understand the behavior of an economy without any reference to money. Money appears to be prevalent in our everyday lives and in economic discussions. We go to work for money, we buy goods with money, we go the bank to borrow money, we hear of vast quantities of money being exchanged in financial markets, we see the value of our money rise and fall relative to other monies, we see the government take our money, and we envy people with gobs of money. Money, money, money. Surely any theory of the way a macroeconomy functions should have money playing a prominent role?

Very few macroeconomists would attempt to argue that money and monetary phenomena are unimportant and undeserving of any attention. The contentious issue among macroeconomists concerns the relative importance of money *vis-à-vis* other factors in determining *real* (as opposed to *nominal*) economic outcomes. The models studied in earlier chapters implicitly took the view that money plays a relatively minor role in determining real outcomes. Since people presumably care about real outcomes, money was ignored. Not all macroeconomists share this view. In the chapters that follow, we will explore their reasons for believing this to be the case.

8.2 What is Money?

When I ask students this question, they often pull out a $5 bill and say 'this is money.' I tell them that is not a very good answer. To demonstrate my point, I pull out a cigarette and say 'no, *this* is money.' Usually, they don't get the point. The point is that anyone can claim that something is money. But what is it about such objects that makes us call them 'money?' What is the definition of money?

Unfortunately, there is no commonly accepted general definition of money. But for our purposes, the following definition appears to be just as good as any other.

Definition: *Money is an object that circulates widely as a means of payment.*

This is not an entirely satisfactory definition, since it does not specify exactly what 'circulate widely' means. Nevertheless, let's work with it anyway. Certainly, the $5 bill seems to fit this definition; hence, it constitutes an *example* of money. But then again, there are alleged cases where cigarettes appear to fit this definition too (e.g., in WWII prisoner-of-war camps). In fact, one could fill a small book describing all the objects that appear to have served as a means of payment throughout the course of human history (e.g., gold, silver, playing cards, shells, stones, animals, etc.)

History is history, one might argue. In modern economies, money is made up of these little paper objects–like this $5 bill. By the way, have you ever looked closely at a (Canadian) $5 bill? It is blue-gold in color featuring a picture of some old guy with a big nose.[1] People actually work hard (sacrifice leisure) to acquire these ugly things? If this is the case, what prevents anyone from printing up similar notes and using them to purchase goods and services? Well, for one thing, such an activity is illegal. If you stared at your $5 bill closely, you may have noticed that it is issued by the *Bank of Canada* and assures us that 'this note is legal tender.' The Bank of Canada is essentially a branch of the federal government. Basically, the government is warning us here that the only type of paper that can legally be used in exchanges is its own paper (and not any other paper that you may be thinking of manufacturing).[2]

But despite the government's monopoly control of the paper money supply, you may be surprised to learn that most of the money supply in any well-developed economy does *not* take the form of small-denomination government-issued paper notes. Most of economy's money supply consists of a particular form of debt instrument issued predominantly by chartered banks in the private

[1] Sir Wilfrid Laurier, Prime Minister of Canada (1896–1911).

[2] The question of *why* governments have legislated themselves such a monopoly is an interesting issue. As usual, there are two (if not more) opposing views. The first view asserts that paper money constitutes a 'natural monopoly,' best handled by a benevolent government. The second view asserts that the government is motivated primarily by the power its monopoly confers in generating revenue.

sector. These debt instruments are called *checkable deposits* or *demand deposit liabilities*.[3] This form of credit does not exist in paper form; i.e., it exists as an electronic book-entry. Your checking account (or checkable saving account) is an example of this type of money. It is called a demand deposit because you have the right to redeem your book-entry money for government cash on demand (as when you visit an ATM to make a cash withdrawal).[4]

Clearly, cash is not the only way in which one may pay for goods and services. Many payments are made using checks or debit cards. When you pay for something using a check, the check instructs the banking system to debit your account by an amount equal to the value of the purchase and credit the merchant's account by the same amount. A debit card transaction does the same thing. No cash need ever change hands in such transactions. Nevertheless, book-entry credits 'circulate' from account to account (instead of hand to hand) and therefore constitute a form of money. Whether money exists in paper or electronic form, its common purpose is to serve as a record-keeping device (i.e., recording the transfer of debits and credits across individuals as they exchange goods and services).

To the extent that the private sector is able to perform such record-keeping services in a relatively efficient manner, one need not be concerned with monetary phenomena when trying to understand the behavior of real economic variables. In a well-function monetary/financial system, the private sector can be expected to create all the money it needs to facilitate economically worthwhile exchanges. The amount of money may fluctuate over time, but it will do so only in response to changes in economic fundamentals. This is the implicit assumption adopted in the first part of the text.

8.3 Private Money

8.3.1 The Neoclassical Model

Let us reconsider the neoclassical model developed in Chapter 2. In particular, refer to Figure 2.4. In that model, we thought of trade between workers and firms as occurring in two stages. In the first stage, workers supply labor to firms in exchange for claims against the output produced by the business sector. We can think of these claims as 'money;' this type of money takes the form of privately-issued liabilities redeemable in output (e.g., gas coupons).

Assume that the business sector creates M_t units of this money at the beginning of date t, where these units are called 'dollars.' This money is used to purchase employment services n_t^*. At this stage, M_t represents nominal GDP

[3] in Canada, for the year 1999, currency in circulation totalled approximately $32 billion (about $1,000 per person), while demand deposits totalled approximately $208 billion.

[4] Historically, chartered banks were allowed to issue small-denomination paper money redeemable on demand for specie (gold or silver coins).

(income-based measure). The nominal wage rate is given by $W_t = M_t/n_t^*$ (the dollars earned per unit of labor).

In the second stage of trading, the household takes its labor income $M_t = W_t n_t^*$ and uses it to purchase the output produced in the business sector, $y_t^* = z_t n_t^*$. At this stage, M_t represents the nominal GDP (expenditure-based). The price-level is given by $P_t = M_t/y_t^*$ (the dollars spent per unit of output).

In this economy, the real wage is given by:

$$w_t^* = \frac{W_t}{P_t} = \frac{M_t}{n_t^*}\frac{z_t n_t^*}{M_t} = z_t.$$

Notice that this is exactly the same real wage that we derived in Chapter 2, where we ignored money. In this economy, people exchange labor for money and money for output. But in an obvious sense, money is just a 'veil' that hides the true nature of the exchange (i.e., the exchange of labor for output).

The other thing we might note from this example is that the economy's nominal variables appear to be *indeterminate*. In other words, we cannot point to any fundamental market forces that might determine (W_t, P_t, M_t). In particular, the business sector can create any amount of money it wants. Having done so, nominal wages and prices simply adjust to the quantity of money in a way that leaves real variables unchanged. Alternatively, we might imagine that the price-level is determined exogenously (say, by social convention). In this case, the nominal wage and the supply of money simply adjusts to leave real variables unchanged. Money and monetary phenomena are simply not important (or interesting) in a neoclassical model.

8.3.2 Wicksell's Triangle: Is Evil the Root of All Money?

Money can be given a more serious role in models that explicitly highlight the 'frictions' that may make money important for facilitating exchange. One such friction is what economists have labelled a 'lack of double coincidence of wants,' or 'lack of double coincidence' (LDC) for short. The basic idea of LDC is as follows. Consider an economy consisting of three people, named A, B and C. Each person has a time-dated endowment y_j, $j = 1, 2, 3$ and a particular preference structure for time-dated consumption. In particular, A has y_3 but wants y_1. B has y_1 but wants y_2. And C has y_2 but wants y_3. Notice that if we pair any two individuals with each other, they will not trade. This is what is meant by a LDC.

But why should we think of these individuals as being paired off? Why don't they all just get together at some central market place (the neoclassical assumption) and swap claims to their output with each other? In particular, imagine that everyone *promises* to deliver to each other the good that they want to consume? If people can be trusted to make good on their promises, then events will unfold as follows. A consumes y_1 in the first period, B consumes

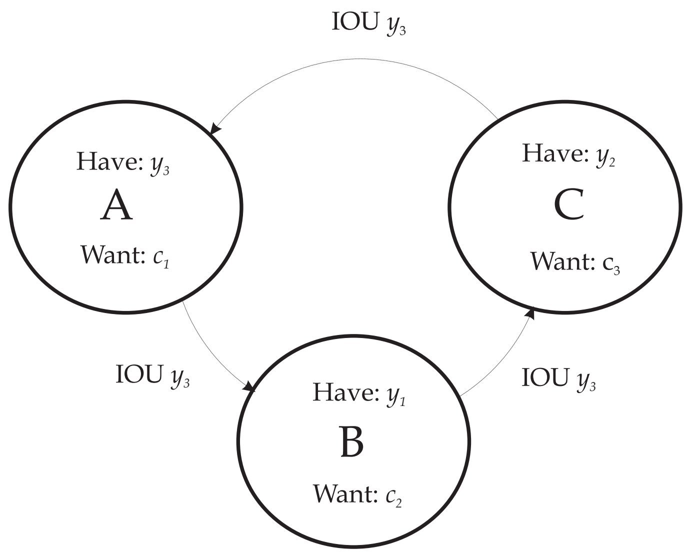
y_2 in the second period, and C consumes y_3 in the third period. Everyone gets what they want (without using money).

As the example above demonstrates, the existence of a LDC is not sufficient to explain the use of money in an economy. We need to introduce some other 'friction.' Kiyotaki and Moore (2000) suggest the following. Imagine that B and C are 'untrustworthy;' i.e., they cannot keep any promises they make. Kiyotaki and Moore describe this lack of commitment on the part of people as an 'evil' that is present in the world. However, imagine that not everyone is 'evil' in this sense. In particular, assume that A is able to keep his promises (and that this is generally known). Clearly, getting together and exchanging promises is not going to work here.

But what will work is the following. Imagine that A issues a debt instrument that represents a claim against y_3. By assumption, A will honor any such claim. In period one, A and B meet, where B gives up his good (y_1) in exchange for A's debt-instrument. This may sound like a strange thing to do from B's perspective; after all, B does not value y_3. Nevertheless, it is rational for B to sacrifice his endowment for such a claim. In particular, while B does not value y_3, he knows that there is somebody out there that does and that this person may have something that B does value. This other person is C. In period two, B and C meet, where C gives up his good (y_2) in exchange for A's debt instrument. It is rational for C to sacrifice his endowment in this way since C knows that he can redeem the debt-instrument for y_3 (which he values). In this exchange pattern, we can clearly identify a monetary instrument (an IOU issued by A). Money is necessary to facilitate exchange in this economy because of a LDC and a general lack of commitment.[5] The trading pattern just described is depicted in Figure 8.1.

[5] Hence, Kiyotaki and Moore are led to proclaim that 'evil is the root of all money.'

Figure 8.1
Wicksell's Triangle



One way to interpret this economy is that A represents a 'bank' (a trusted financial institution), while B and C individuals represent various anonymous members of the population.[6] In the first period, B makes a 'deposit' (consisting of y_1) and A issues a liability (in exchange for the deposit) that is collateralized by the bank's 'capital' (a claim to future output, y_3). This collateralized private debt instrument then circulates in a series of exchanges, before it is ultimately redeemed at the bank. If one is willing to interpret y_3 as 'gold,' then this economy resembles the way money was created in many historical episodes (e.g., the so-called free-banking era in the United States, 1836–1863).

Note that in this model, money plays an important role in facilitating transactions (without money, there would be no trade). Furthermore, observe that the trading patterns that do occur coincide with what would have occurred in a neoclassical model with well-functioning markets (i.e., where all agents can commit to their promises). Here then, is another example of a general point that I made already: To the extent that money/financial markets work well, we can ignore money if we are only interested in understanding the behavior of real allocations.

[6] If individuals view themselves as anonymous, then they will not be willing to accept each other's promises.

8.3.3 Government Money

Traditionally, most macroeconomic models that include money view things from an opposite extreme; i.e., under the assumption that the economy's money supply is created (or at least controlled) by the government (instead of the private sector). In the context of the models described above, this change in perspective turns out to be innocuous.

Consider, for example, the neoclassical labor market model. Here we can imagine that the government creates M_t dollars of money and either lends or gives it to the business sector. The business sector then uses this money to purchase labor and households subsequently use their money income to purchase output. Since M_t is viewed as exogenous, the equilibrium price-level is given by:

$$P_t^* = \frac{M_t}{y_t^*}; \tag{8.1}$$

with the nominal wage being determined by $W_t^* = z_t/P_t^*$.

Likewise, in the context of the Wicksellian model, imagine that the government creates M units of money and then lends it to the A individual. The sequence of exchanges may now proceed on a sequence of spot markets where goods trade for government money. The equilibrium price-level at each date is given by $P_t^* = M/y_t$.

The only 'benefit' of viewing things in this manner is that equilibrium nominal variables are now determinate (by assuming that M_t is exogenous). In particular, the price-level (and inflation) is now under control of government policy. Whether this view holds in reality, however, depends on the extent to which government policy may actually control the economy's money supply. As I mentioned earlier, in reality, most of an economy's supply (i.e., that part that does not constitute paper) is created by the private sector.

8.4 The Quantity Theory of Money

One of the oldest theories in macroeconomics is the so-called *Quantity Theory of Money* (QTM).[7] The QTM is perhaps an inaccurate name for the theory, since it does not constitute a theory of money. In fact, the QTM constitutes a theory of the *price-level* (and inflation), which happens to emphasize the role played in such matters by an exogenously determined quantity of money.[8]

The QTM begins with the *assumption* that individuals demand money.[9] In

[7] See, for example, Laidler (1985), for an extensive discussion of the QTM.

[8] As such, the QTM may be more accurately labelled as 'The Quantity of Money Theory of the Price-level.'

[9] Hence, the QTM is not a theory of money since it does not specify the conditions under which people value money. Note that this is in contrast to the Wicksellian model studied earlier.

its simplest version, the theory assumes that the demand for money is proportional to the volume of trade, say, as measured by the nominal GDP; i.e.,

$$M_t^D = \kappa P_t y_t, \tag{8.2}$$

where $\kappa > 0$ is an exogenous parameter and y_t is viewed as being determined by a neoclassical labor market (as in Chapter 2).

The QTM then asserts that the supply of money fully controlled by government policy; i.e.,

$$M_t^S = M_t. \tag{8.3}$$

The equilibrium price-level is then viewed as being determined by a 'money market clearing condition;' i.e., combining equations (8.2) and 8.3):

$$M_t = \kappa P_t y_t;$$

which can be solved for:

$$P_t^* = \frac{M_t}{\kappa y_t}. \tag{8.4}$$

Note the resemblance between (8.4) and (8.1).

Hence, the QTM predicts that the price-level is proportional to the stock of money. In particular, for a given demand for real money balances (κy_t), and exogenous increase in M_t will lead to an increase in the price-level. Intuitively, if there is more money 'chasing' a given level of output, one would expect the price of output (measured in units of money) to increase.

- **Exercise 8.1**. The price-level measures the value of output in units of money. Alternatively, one could think of a value for money measured in units of output. According to (8.4), what happens to the value of money when M_t is increased? Explain.

Note that since the real GDP is determined independently of M_t, exogenous changes in the money supply have no effect of on any real economic variables. Monetary policy here can only influence nominal variables. Any model that features this property is said to feature *money neutrality.*

Definition: Money is said to be *neutral* if exogenous changes in the supply of money have no effect on real quantities and real prices.

If money is indeed neutral in reality, then there would be little point in studying money or monetary policy. Most economists take the view that money is neutral in the 'long-run,' but not necessarily in the 'short-run.' The evidence for short-run neutrality, however, does not appear to hold universally; e.g., see Appendix 8.A.

With a little more work, the QTM can be transformed into a theory of inflation. Inflation is defined as the *rate of growth* of the price-level; i.e.,

$$\pi_t \equiv \frac{P_t - P_{t-1}}{P_{t-1}}.$$

Note that π_t may be either positive or negative. A negative inflation rate is called *deflation*. Note that we can alternatively define the gross rate of inflation as $\Pi_t \equiv P_t/P_{t-1} = (1 + \pi_t)$

- **Exercise 8.2**. What is happening to the value of money over time when an economy is experiencing an inflation/deflation?

Assume that the government expands the supply of money at some constant rate μ, so that $M_t = (1 + \mu)M_{t-1}$. You can think of M_{t-1} as being 'old' money and μM_{t-1} as representing 'new' money that injected into the economy in some manner (the QTM does not specify any details concerning how this new money is injected). Furthermore, assume that the 'real' economy grows at some exogenous rate γ, so that $y_t = (1 + \gamma)y_{t-1.}$ Now, from (8.4) one can derive:

$$\frac{P_t^*}{P_{t-1}^*} = \frac{M_t}{M_{t-1}} \frac{y_{t-1}}{y_t}$$
$$\Pi^* = \frac{(1+\mu)}{(1+\gamma)}.$$

Note that if μ and γ are 'small' numbers, then $\pi^* \approx \mu - \gamma$.

Thus, the QTM predicts that the economy's inflation rate depends positively on the *rate of growth* of the money supply and negatively on the *rate of growth* of the real economy (i.e., the rate of growth of the demand for real money balances). Again, the intuition is simple. If the money supply grows faster than the demand for real money balances, then the price of output will rise (the value of money will fall) over time.

In this version of the QTM, the monetary policy parameter is given by μ. Note that since the 'real' economy here (i.e., y_t and γ) is viewed as being determined independently of the growth rate of money, exogenous changes in μ have no effect on any real variables. Any model that features this property is said to feature *money superneutrality*.

Definition: Money is said to be *superneutral* if exogenous changes in the *rate of growth* of the supply of money have no effect on real quantities and real prices.

The evidence for superneutrality is mixed. Most economists take the view that money is superneutral, at least, for low to moderate inflations/deflations. On the other hand, historical episodes of extremely high inflations (hyperinflations) seem to suggest that money is not superneutral. When inflation is very high, the value of money declines very rapidly, inducing people to take extraordinary measures (involving real resource costs) to economize on their money holdings. Over the period July-November 1923 in Germany, for example, the price-level rose by 854,000,000,000%. According to some sources:

> "Workmen are given their pay twice a day now–in the morning and in the afternoon, with a recess of a half-hour each time so that they can rush out and buy things–for if they waited a few hours the value of their money would drop so far that their children would not get half enough food to feel satisfied."

Evidently, merchants eventually found that they had trouble marking up their prices as fast enough.

> "So they left the price marks as they were and posted (hourly) a new multiplication factor. The actual price marked on the goods had to be multiplied by this factor to determine the price which had to be paid for the goods. Every hour the merchant would call up the bank and receive the latest quotation upon the dollar. He would then alter his multiplication factor to suit and would perhaps add a bit in anticipation of the next quotation. Banks had whole batteries of telephone boys who answered each call as follows: '100 milliarden, bitte sehr, guten Tag.' Which meant: 'The present quotation on the dollar is 100 billion marks, thank you, good day."[10]

According to the QTM, episodes like the German hyperinflation are 'caused' by an overly expansionary monetary policy. High money growth rates imply high inflation. The way to prevent inflation is keep the money supply expanding at a moderate rate (approximately equal to the growth rate of the real economy). Indeed, if one looks at a cross-section of countries, the correlation between inflation and money growth appears to be very high. The same is true for time-series observations within a country over 'long' periods of time (the correlation is not as strong over 'short' intervals of time). This type of evidence is usually interpreted as lending support for the QTM. On the other hand, Smith (1985) documents one of several historical episodes in which rapid money supply growth appears to have resulted in little, if any, inflation.

In any case, even if there is a strong positive correlation between inflation and money growth, care must be taken in inferring a particular direction of causality. The QTM asserts that inflation is 'caused' by monetary policy. One way to think about this is that some exogenous event increases a government's demand for resources (e.g., the need to finance post WWI war reparations, in the case of Germany) and the way it chooses to finance this need is by creating new money.

[10] These quotes were obtained from: http:// ingrimayne.saintjoe.edu/ econ/ Economic-Catastrophe/ HyperInflation.html

Alternatively, one might take the view that the direction of causality works in reverse. There appears to be a hint of this in the previous quote which suggests that merchants increased their product prices in anticipation of the future value of money. One way in which this might happen is through a 'wage-price spiral' that is accommodated by the government. That is, instead of assuming that M_t is chosen exogenously, imagine that the government prints an amount of money that is demanded by the private sector. In the context of our simple neoclassical labor market model, the amount of money printed (in the first stage) will depend on the nominal wage; i.e., $M_t^* = W_t n_t^*$. Now, imagine that the nominal wage is chosen in a way that targets the equilibrium real wage z_t; i.e., $W_t = z_t P_t^e$, where P_t^e denotes the price-level that is expected to occur (in the second stage). In this setup, the rate of money growth is determined by the expected rate of inflation; i.e.,

$$\frac{M_t^*}{M_{t-1}^*} = \frac{P_t^e}{P_{t-1}^e}.$$

If these expectations are correct, then the *actual* inflation rate will correspond to the *expected* inflation rate.

A wage-price spiral may be initiated then by an exogenous increase in inflation expectations. Higher expectations of inflation lead workers to negotiate higher nominal wages (to maintain their real wages). The business sector responds by either creating or acquiring the necessary money to accommodate these wage demands. The additional money created in this wage then generates a higher inflation (confirming expectations).

An economist trained in the QTM is likely to accept these logical possibilities. However, he or she would nevertheless maintain that inflation is 'always and everywhere a monetary phenomenon.' In particular, while the German hyperinflation may have been 'caused' by the government's revenue needs, an independent monetary authority could have prevented the hyperinflation by refusing to accommodate the demands of the fiscal authority. Likewise, a wage-price spiral can be avoided by having a 'strong' monetary authority that is unwilling to accommodate the private sector's (expectations driven) demand for money.

8.5 The Nominal Interest Rate

In earlier chapters, we introduced the concept of a real interest rate as a relative price of time-dated output and discussed how the equilibrium real interest rate is determined in a neoclassical model; i.e., see Sections 4.5 and 6.5. In reality, there are rarely any direct measures of the real interest rate. Most interest rates that are quoted are nominal. The nominal interest rate is also a relative price; it is the relative price of time-dated money.

To examine the link between the real and nominal interest rate, consider the following two debt instruments. Imagine that the government issues two types of bonds: a *nominal bond* (by far the most common) and a *real bond* (considerably more rare). Assume that both types of bond instruments are risk-free. A nominal bond constitutes a contract stipulated in nominal terms. For example, if I purchase a nominal bond for B_t dollars at some date t, the government promises to return $R_t^n B_t$ dollars (principal and interest) at some future date $t+1$. Here, R_t^n denotes the (gross) nominal interest rate. The nominal interest rate tells us that one dollar today is worth $1/R_t^n$ dollars in the future.

Similarly, a real bond constitutes a contract stipulated in real terms. For example, if I purchase a real bond for b_t units of output at some date t, the government promises to return $R_t b_t$ units of output (principal and interest) at some future date $t+1$. Here, R_t denotes the (gross) real interest rate. The real interest rate tells us that one unit of output today is worth $1/R_t$ units of output in the future.

In practice, the contractual stipulations in a real bond are also specified in units of money. In addition, however, the contract links the dollar repayment amount to the future price-level; i.e., P_{t+1}. In other words, the difference between a nominal bond and a real bond is that the latter is indexed to inflation.

Thus, if I give up B_t dollars today to purchase either a real or nominal bond, I am in effect sacrificing $B_t/P_t = b_t$ units of output (which I could have purchased and consumed). A nominal bond returns $R_t^n B_t$ dollars to me in the future. The purchasing power of this future money is given by $R_t^n B_t/P_{t+1}$. A real bond returns $R_t b_t$ units of output (purchasing power) to me in the future. Now let us compare the *real* rates of return on each of these debt instruments.

The rate of return on an asset is defined as:

$$ROR \equiv \frac{\text{Return}}{\text{Cost}}.$$

Hence, the *real* rate of return on a nominal bond is given by:

$$ROR_{\text{nominal bond}} = \frac{R_t^n B_t/P_{t+1}}{B_t/P_t} = \frac{R_t^n}{\Pi_t}.$$

The real rate of return on a real bond is given by:

$$ROR_{\text{real bond}} = \frac{R_t b_t}{b_t} = R_t.$$

Which of these two assets would you rather invest in? Recall that both debt instruments are free of risk. If this is the case, you should prefer to invest in the bond instrument that yields the higher real return (the nominal return is irrelevant). In fact, for both of these bonds to be willing held in the wealth portfolios of individuals, it must be the case that the two bonds earn the *same* real return; i.e.,

$$R_t = \frac{R_t^n}{\Pi_t}. \tag{8.5}$$

This condition is constitutes a simple application of a *no-arbitrage-condition*. If this condition did not hold, then bond traders would be able to make huge amounts of profit, for example, by shorting the lower return instrument and using the proceeds to purchase long positions in the higher return instrument. Such arbitrage opportunities are not likely to last very long in a competitive financial market. The sell pressure on the low return bond will reduce its price, thereby increasing its yield. Likewise, the buy pressure on the high return bond will lower its price, thereby increasing its yield. In equilibrium, arbitrage opportunities like this will cease to exist; i.e., the returns must adjust to satisfy (8.5).

8.5.1 The Fisher Equation

Condition (8.5) can be rewritten as:

$$R_t^n = R_t \Pi_t;$$

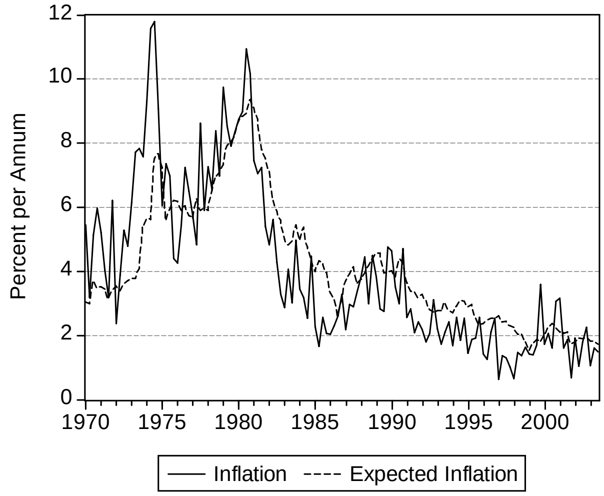
or, in terms of net rates:

$$r_t^n \approx r_t + \pi_t.$$

Written in this way, this condition is often referred to as the *Fisher equation*. The Fisher equation constitutes a theory of the nominal interest rate. It claims that the (net) nominal interest rate should be approximately equal to the (net) real interest rate plus the (net) rate of inflation. The intuition is simple. Given that there are other assets (e.g., capital or indexed bonds) in the economy that yield a real return r_t, the nominal return on a nominal bond had better return enough future dollars to compensate for the expected loss in the purchasing power of money (inflation). Only if the nominal interest rate is high enough to compensate for (expected) inflation will individuals be willing to hold an non-indexed nominal bond.

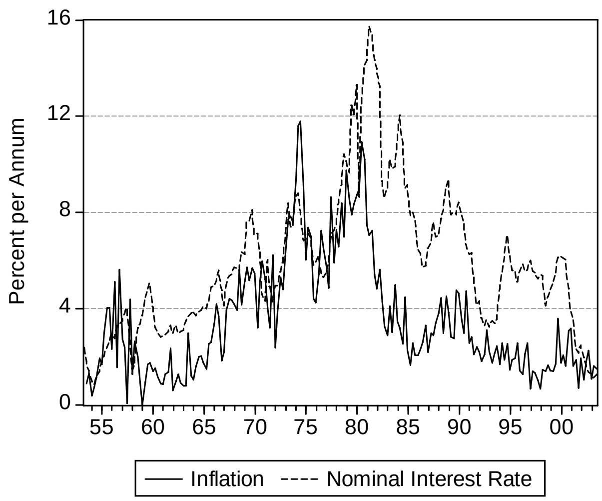
Evaluating the empirical legitimacy of the Fisher equation is not a straightforward exercise. For one thing, properly stated, the theory suggests that the nominal interest rate should be a function of the *expected* real interest rate and the *expected* rate of inflation. Direct measures of such expectations can be hard to come by (especially of the former). Often what is done is to assume that the expected inflation rate more or less tracks the actual inflation rate, at least, over long periods of time. According to the Livingston Survey of inflation expectations, this is probably not a bad assumption, although there does appear to be a tendency for expectations to lag actual movements in inflation; i.e., see Figure 8.2.

FIGURE 8.2
Inflation and Expected Inflation
United States 1970.1 - 2003.3



Consider next the time-series behavior of the nominal interest rate and inflation in the United States:

FIGURE 8.3
Inflation and the Nominal Interest Rate
United States 1953.2 - 2003.3



According to Figure 8.3, the long-term movements in the nominal interest rate do appear to follow at least the trend movements in inflation (and hence, inflation expectations) in a manner consistent with the Fisher equation. Note, however, that the correlation is not perfect, especially for short-run movements. This latter observation is not necessarily inconsistent with the Fisher equation since these short-run movements could be the result of movements in the (short-run) real interest rate. In fact, because the logic of the Fisher equation is viewed as so compelling, economists typically assume that it is true and then use the equation to derive a measure of the real interest rate!

8.6 A Rate of Return Dominance Puzzle

Let us reconsider the no-arbitrage principle (NAC) discussed earlier in reference to the Fisher equation. This principle can be formally stated as follows:

No-Arbitrage Principle: *Any two assets sharing identical risk characteristics must yield the same expected return if they are both to be held willingly in the wealth portfolios of individuals.*

Stated another way, if one of these two assets does yield a lower rate of return, then it will be driven out of existence. Among economists, the no-arbitrage principle has essentially attained the status of religion. There is a good reason for this. In particular, the idea that unexploited riskless profit opportunities exist for any relevant length of time seems almost impossible to imagine.

Now let us consider the following two assets, both of which are issued by the government. One asset is called a *bond*, and the other is called *money*. A bond represents a claim against future money. But then, money also represents a claim against future money. If I hold B dollars of one-year government bonds, at the end of the year these bonds are transformed into $R^n B$ dollars. If instead I hold M dollars of government money, at the end of the year this money is 'transformed' into M dollars (since paper money does not pay interest). In other words, government money is just another type of government bond; i.e., it is a bond that pays zero nominal interest.

What is interesting about this example is that it appears (on the surface at least) to violate the no-arbitrage principal (at least, assuming that government bonds are free of nominal risk). Why do people choose to hold government money when money is so obviously dominated in rate of return? Are individuals irrational? Why is this rate of return differential not arbitraged away? Alternatively, why do government bonds not drive government money out of circulation?

The explanations for this apparent violation of the no-arbitrage principle fall under two categories. The first category is one that you've probably thought of already. The argument goes something like this. Government money is a 'special' type of asset. In particular, it is a 'liquid' asset, whereas a government bond is not. For example, just try buying a cup of coffee (or anything else) with a government bond. Thus, while the *pecuniary* (i.e., monetary) return on money may be low, money confers a *non-pecuniary* return in the form of 'liquidity' services. Thus, observing differences in the *pecuniary* rates of return between money and bonds is not necessarily a violation of the no-arbitrage principle; i.e., the apparent gap between these two returns may simply reflect the *non-pecuniary* return on money.

The argument just stated sounds compelling enough to most people. But upon further examination, it appears unsatisfactory. In particular, the explanation simply asserts that government money is a 'special' asset without explaining why this might be the case. It does refer to the idea that money is 'liquid,' but fails to define the term or explain what it is about money that makes it 'liquid.' Furthermore, it is not at all apparent that such a rate of return differential could not be arbitraged away by the banking system. For example, a bank should, in principle, be able to purchase a government bond and then create its own paper money 'backed' by such an instrument. Banks could make huge profits by printing zero interest paper while earning interest on the bond it holds in reserve. Competition among banks would then either compel them to pay interest on their money, or drive the interest on bonds to zero.[11]

[11] A small interest rate differential may remain reflecting the cost of intermediation.

You might object to this argument on the ground that while the idea sounds good in principle, in practice banks are legally prevented from issuing their own paper money (since 1935 in Canada). Good point. In fact, such a point represents the *legal restrictions hypothesis* for why government money is dominated in rate of return (Wallace, 1983).

So now you agree that there is nothing particularly 'special' about government paper money. Private banks can issue paper money too (and have done so in the past). What prevents banks from doing so today is largely the product of a legal restriction (i.e., the government wishes to maintain a monopoly over the paper money supply). Government bonds are not useful for payments because they are either: [1] issued in very large denominations (e.g., $10,000 or more); or [2] they exist only as electronic book-entries (as is mainly the case these days). Thus, the no-arbitrage principle is not violated because the principle only holds in the absence of government trade restrictions.

As a corollary, the legal restrictions hypothesis predicts that the rate of return differential between money and bonds would disappear if one of the following two government reforms were implemented. First, if the government (in particular, the treasury or finance department) began to issue paper bonds in the full range of denominations offered by the central bank. Second, if the government was to alter legislation that prevented banks (or any other private agency) from issuing its own paper money.[12]

8.6.1 The Friedman Rule

Is inflation/deflation 'good' or 'bad' for the economy? While we have not, as of yet, developed a model that is capable of examining the welfare implications of inflation, it is nevertheless useful at this stage to ponder the question for what lies ahead.

An extremely robust result in most economic models is that economic efficiency (in the sense of Pareto optimality) requires that no-arbitrage conditions be satisfied. Let us consider the *real* rates of return on two types of assets: government money and risk-free capital (if such a thing exists).[13] The real return on capital is R. The real return on money is $1/\Pi$ (since money is like a zero interest nominal bond). The no-arbitrage principle then asserts that efficiency requires:

$$\frac{1}{\Pi} = R. \tag{8.6}$$

Equation (8.6) is the celebrated *Friedman rule*. Recall from the Fisher equa-

[12] While either reform is likely to generate rate of return equality between money and bonds, we cannot say (without further analysis) whether the nominal return will be positive or zero.

[13] The demand deposit liabilities of modern-day chartered banks perhaps constitute an example.

tion (8.5) that the nominal interest rate is given by $R^n = R\Pi$. Hence, the Friedman rule is asserting that an optimal monetary policy should operate in a manner that drives the (net) nominal interest rate to zero; i.e., $R\Pi = 1$. If $R > 1$ (as is normally the case), then this policy recommends engineering a *deflation*; i.e., $\Pi = 1/R < 1$. If $R < 1$ (as may be the case in present day Japan), then this policy recommends engineering an *inflation*; i.e., $\Pi = 1/R > 1$. Price-level stability (zero inflation) is only recommended when the (net) real interest rate is zero.

Since the Friedman rule is based on a no-arbitrage principle, it is difficult to dispute it's logic. Nevertheless, almost no one in policy circles takes the Friedman rule seriously. Central bankers, in particular, appear to be highly averse to the idea of a zero nominal interest rate. The reasons for why this might be the case will be explored in a later chapter. But for now, we must simply regard any departure from the Friedman rule as an unresolved 'puzzle.'

8.7 Inflation Uncertainty

If inflation was always easily forecastable, then it is hard to imagine how (at least moderate) inflations or deflations may pose a pressing economic problem (at least, relative to all the other things we have to worry about). Nominal prices could in this case be contractually agreed upon in a way that leaves the underlying 'real' prices (including wages and interest rates) at their 'correct' levels.

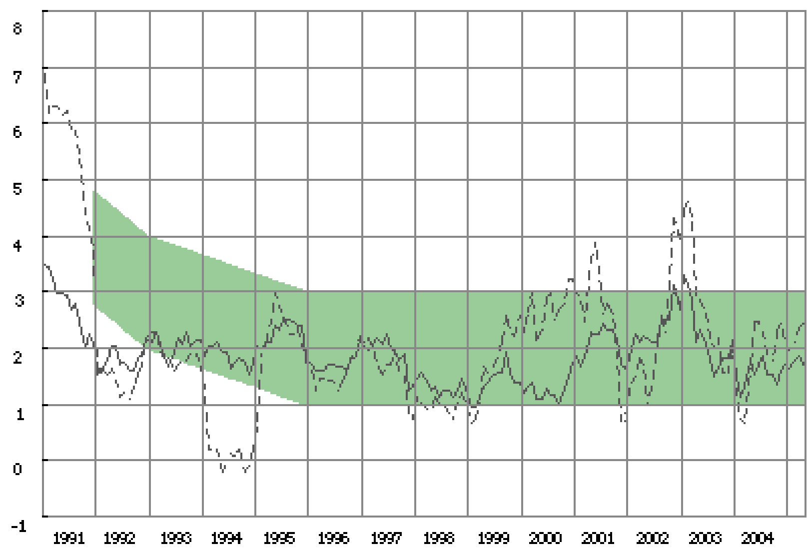
Of course, inflation is not always easily forecastable. This appears to be especially true for economies experiencing very high rates of inflation. It is a fact of life that most real-world contracts are stated in nominal terms and that these terms depend, at least in part, on the forecast of inflation. If inflation is highly variable, it is not easy for nominal contracts to ensure the 'proper' allocation of real resources. Unexpected inflation is viewed as being undesirable for two reasons. First, if contracts are not indexed to inflation (normally, they are not) and if contracts are costly to renegotiate (as is surely the case), then an unexpected inflation results in a redistribution of resources (for example, from creditors to debtors). Second, if indexation and/or renegotiation is costly, then inflation uncertainty is likely to entail resource costs and the curtailment of economic activity.

These are the primary reasons for why it is a stated policy of many central banks to keep inflation 'low and stable.'[14] To this end, many central banks have adopted an *inflation target*. The Bank of Canada, for example, has (since 1991) adopted an inflation target of 2% (not the Friedman rule!) with an operating band of plus/minus 1%; i.e., see Figure 8.4. The general consensus appears to be that inflation targets work well toward the goal of keeping inflation 'low and

[14] See, for example, www.bankofcanada.ca/ en/ inside.htm

stable.'[15]

Figure 8.4
Bank of Canada Inflation Target



8.8 Summary

Money is an asset whose role is to record individual transactions. In its role as a record-keeping device, money serves to facilitate exchange, and hence improve economic welfare.

Money exists in two basic forms: small denomination paper and electronic book-entry. In most modern economies, the government (via a central bank) maintains monopoly control over the supply of small denomination paper, while the private sector (via the banking system) is left to determine the supply of book-entry money. Since the vast majority of money is in the form of book-entry, it is not clear to what extent a government can control the total supply of money (the sum of paper and book-entry money). In practice, however, various legal restrictions on the banking sector likely imply that the government can exert some influence on the supply of book-entry money (and hence, the total money supply).

To understand the behavior of nominal variables, one must have a theory that includes some role for money. But since economic welfare depends ultimately on real variables, the study of money (and monetary policy) is only

[15] The interested reader can refer to Bernanke, Laubach, Mishkin and Posen (1999).

relevant to the extent it influences real economic activity. In the neoclassical model, money may be important for economic efficiency, but money itself is not a source of economic disturbance nor do monetary factors influence the way an economy responds to other shocks.

The neoclassical view of money may provide a good approximation for some historical episodes in which the money supply was almost entirely provided by a relatively free and competitive banking system (e.g., the Scottish and U.S. 'free-banking' eras). There is, however, a considerable ongoing debate of this issue. In any case, in most modern (and historical) economies, the government exerts at least some control over money supply. Furthermore, various 'contracting frictions' may be severe enough to render money non-neutral, so that 'shocks' to monetary policy may potentially constitute an important source of the business cycle. Even in the absence of monetary policy shocks, these 'frictions' may influence the way an economy responds to other types of disturbances.

8.9 Problems

1. Consider the Wicksellian model in Figure 8.1. One way to imagine trade taking place is as follows. First, B makes a 'gift' to A. Second, C makes a 'gift' to B. Finally, A makes a 'gift' to C. If society could keep a complete record of such gifts, would there be any role for money? Explain how money can be thought of as a substitute for such a public record-keeping technology.

2. We are often counselled by financial planners to set aside at least a part of our saving in the form of 'safe' government bonds and money (cash). Explain why money and bonds are not risk-free debt instruments in the future rate of inflation is uncertain.

3. Many macroeconomic textbooks make reference to the notion of a 'monetary policy shock.' Does this concept make any sense to you? Why, in particular, would a monetary authority want to 'shock' the economy? Or is there some other way of interpreting such a shock? Discuss.

8.10 References

1. Bernanke, Ben S, Thomas Laubach, Frederic S. Miskin and Adam S. Posen (1999). *Inflation Targeting: Lessons from the International Experience*, Princeton University Press, Princeton, New Jersey.

2. Laidler, David E. W. (1985). *The Demand for Money*, Harper & Row, Publishers, New York.

3. Smith, Bruce D. (1985). "American Colonial Monetary Regimes: The Failure of the Quantity Theory and Some Evidence in Favour of an Alternative View," *Canadian Journal of Economics*, XVIII(3): 531–565.

4. Wallace, Neil (1983). "A Legal Restrictions Theory of the Demand for 'Money' and the Role of Monetary Policy," *Federal Reserve Bank of Minneapolis Quarterly Review*, 7(1): 1–7.

Chapter 9

The New-Keynesian View

9.1 Introduction

Many economists and policymakers do not believe that money is neutral, at least, in the 'short-run.' To take this view, one must believe the following two things: [1] that the economy's relevant money supply is determined (or at least, greatly influenced) by policy; and [2] that 'contracting frictions' are present that make at least some nominal variables 'sticky;' at least, for 'short, but sufficiently long' periods of time. The term 'sticky' is meant to capture the idea that some nominal variables do not react 'immediately' to shocks (in particular, monetary policy shocks). A prominent and influential strand of the literature that emphasizes the importance of sticky nominal prices operates under the label 'New-Keynesian' economics.[1]

9.2 Money Non-Neutrality

New Keynesian models typically feature either sticky prices or sticky wages, but not both. I am not sure what accounts for this either or treatment. Perhaps it is because if both prices and wages are sticky, then real wages would be sticky, and generating money non-neutrality may be more difficult. In what follows, I consider a model with sticky wages.

[1] The label 'New-Keynesian' is somewhat ironic in light of the fact that Keynes (1936) appeared to take the view that nominal prices 'too' flexible and destabilizing. For example, a rapid decline in product prices might lead to a ruinous 'debt-deflation' cycle (as falling prices would increase real debt burdens). Likewise, rapidly falling nominal wages contribute to a decline in demand that would exacerbate an economic downturn.

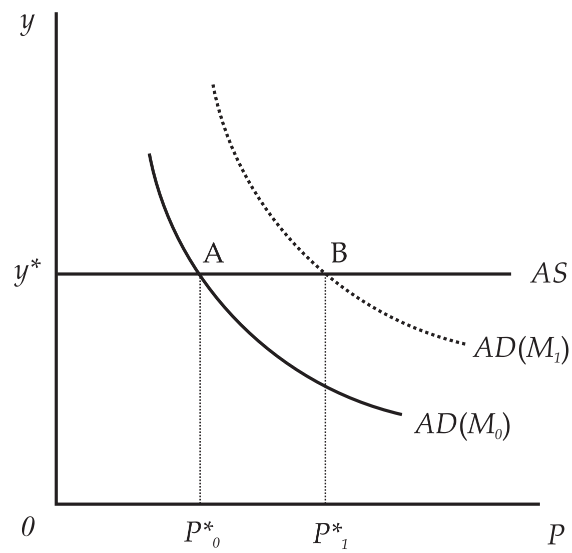
9.2.1 A Basic Neoclassical Model

To begin, consider the neoclassical model of the labor market, for example, as developed in Appendix 2.A. Profit maximization there implies a downward sloping labor demand function $n^D(w)$, where w denotes the real wage. Utility maximization on the part of households implies an upward sloping labor supply function (assuming that the substitution effect dominates the wealth effect for real wage changes) $n^S(w)$. In a neoclassical equilibrium, the equilibrium real wage and employment are determined by $n^S(w^*) = n^D(w^*) = n^*$. This level of employment generates a 'natural' level of real GDP; $y^* = F(n^*)$.

Now, to introduce money into the model, let us appeal to the *Quantity Theory of Money*, which asserts that for a given money supply M, the equilibrium price-level is determined by $P^* = M/y^*$; i.e., see Chapter 8. The equilibrium nominal wage is then given by $W^* = w^*P^*$.

Figure 9.1 depicts the neoclassical equilibrium graphically. The figure depicts an 'aggregate supply' (AS) function and an 'aggregate demand' (AD) function. These labels are perhaps not the best ones available, since these 'supply' and 'demand' functions do not correspond to standard microeconomic definitions. The way to think of the AS curve is that it represents all the output-price combinations that are consistent with equilibrium in the labor market. Since equilibrium in the labor market does not depend on the price-level, the AS curve is horizontal. The way to think of the AD curve is that it represents all the output-price combinations that are consistent with equilibrium in the money market (for a given level of M). The AD curve slopes downward from left to right because a higher price-level reduces the supply of real money balances, which implies that a lower level of output is need to clear the money market. When the money supply is equal to M_0, the general equilibrium occurs at point A, where *both* the labor and money market are in equilibrium. An exogenous increase in the money supply to $M_1 > M_0$ moves the economy to point B, leaving all real variables unchanged. In other words, money is neutral.

FIGURE 9.1
Response to a Money Shock: Neoclassical Model



9.2.2 A Basic Keynesian Model

From a New-Keynesian perspective, point A in Figure 9.1 represents how an economy might be expected to behave in the 'long-run' (i.e., the amount of time it takes nominal prices and wages to adjust to their equilibrium levels). In the 'short-run,' however, the economy may react quite differently to a money supply shock.

Imagine, for example, that workers negotiate a nominal wage contract of the following form. Workers agree to supply all the labor that is demanded from them at some given nominal wage W. The nominal wage is not indexed to the price-level. Nor can this nominal wage be renegotiated in the 'short-run.' The lack of indexation or renegotiation poses no problem if the price-level remains constant over time. Perhaps this is one way to rationalize a non-indexed wage. That is, if workers expect the price-level to remain relatively stable over time, there is no sense in negotiating (costly) indexation clauses into their wage contract.

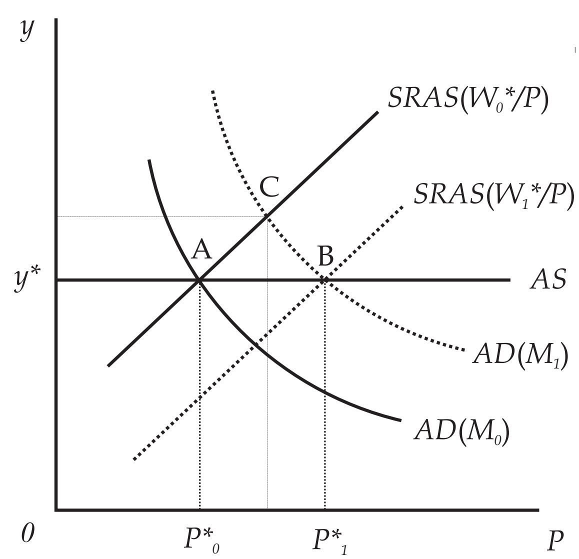
With labor supply modeled in this way, the level of employment is determined solely by labor demand; i.e., $n^D(w)$, where $w = W/P$. Notice that for a fixed nominal wage W, the demand for labor is *increasing* in the price-level. This is because a higher price-level reduces the real wage, and hence, the real cost of labor to firms. Consequently, it follows that the AS curve is–in the 'short-run' at least–an *increasing* function of the price-level; i.e., $y^S(W/P) = F\left[n^D(W/P)\right]$.

Let us label this relationship as the SRAS (short-run aggregate supply) function.

The AD relationship is as before. Together the SRAS and the AD curve are plotted in Figure 9.2. Assume that the wage is initially fixed at it's neoclassical level; i.e., $W = W^*$. In this case, the equilibrium is given by point A. At this point, the economy is said to be both in a 'short' and 'long' run equilibrium.

Now, imagine that the money supply is unexpectedly increased (for some unexplained reason). Then the AD curve shifts 'up' as in Figure 9.1. As in the neoclassical model, the effect of this shock is to put upward pressure on the price-level. However, unlike the neoclassical model, we see here that the level of output (and employment) rises as well; i.e., in the 'short-run,' the economy moves to point C in Figure 9.2. The level of output increases because the expansion in money supply ultimately reduces real labor costs (since the nominal wage is fixed and since prices are higher). In other words, money is not neutral–at least, in the 'short-run.' In the 'long-run,' one can imagine that workers would react to this development by demanding higher wages (i.e., $W_1^* > W_0^*$). As this process unfolds, the SRAS shifts back 'down,' and the economy eventually moves to point B.

FIGURE 9.2
Response to a Money Shock: Keynesian Model



- **Exercise 9.1**. Consider Figure 9.2 and imagine that the economy is initially at point A. Now, imagine that the economy experiences a positive productivity shock. The effect of this shock is to shift the AS and SRAS 'up' by the same distance.

(a) Assuming that M remains fixed, explain how the economy reacts both in the short and long run (it will be helpful to first work through the neoclassical case).

(b) How might a monetary authority react to such a shock to facilitate the transition to the higher long-run level of output? What implications would such a policy have for the price-level? Is such a policy likely to improve welfare? Explain.

9.3 The IS-LM-FE Model

In order to set up the discussion that follows in the next section, it will be useful to present an extension to the basic Keynesian model developed above. The extension involves introducing a role for the interest rate. As we know from earlier chapters, an interest rate is an intertemporal price. Proper modeling of the interest rate should entail an explicit description of the economy's dynamic structure. But developing an explicitly dynamic model with sticky prices involves a complicated analysis. For this reason, we follow convention and employ a number of 'short-cuts' by favoring intuitive arguments over rigorous derivation. The basic intuition, however, will survive a more rigorous theoretical treatment.

The extended model is called the IS-LM-FE model and is essentially an extension of the basic Keynesian AS-AD model developed above. The IS-LM-FE version highlights the relationship between output and the interest rate. The extended AS-AD version highlights the relationship between output and the price-level. Both versions constitute the *same* model presented graphically in different spaces; i.e., (y, R) space versus (y, P) space.

In what follows, we will take R to denote both the real and nominal interest rate. From the Fisher equation, we know that this will only be true if expected inflation is zero. All that we really need, however, is to assume that inflation expectations are 'sticky' in the short-run. But for simplicity, it is assumed here that inflation expectations are fixed at zero.

9.3.1 The FE Curve

FE stands for 'Full Employment.' The FE curve is defined as the combination of (y, R) that are consistent with equilibrium in the labor market *for a given* P. If intertemporal market forces are relatively weak as far as the current labor market is concerned, then we can essentially stick to the 'static' labor market model developed above. In this case, equilibrium in the labor market is independent of (among other things) the interest rate. Hence, if we draw a graph with y on the y-axis and R on the x-axis, the FE curve is horizontal. That is, the FE curve consists of the the (y, R) combinations that satisfy:

$$y^* = F(n^*); \tag{9.1}$$

where n^* is determined by $n^S(w^*) = n^D(w^*) = n^*$.

If the nominal wage is sticky, however, there is also a 'short-run' FE curve (SRFE) that will depend on W and P. This is just the analog of the SRAS curve discussed earlier. For a fixed W, an increase in P will shift the SRFE curve 'up;' and a decrease in P will shift the SRFE curve 'down.' All this tells us is that in the short-run, the supply of output is increasing in the price-level; i.e., the SRAS curve is an increasing function of P. The SRFE curve consists of all the (y, R) combinations that satisfy:

$$y^S(W/P) = F\left[n^D(W/P)\right]. \tag{9.2}$$

Note that the SRFE curve does not depend on R (although, its position will shift with changes in W or P that alter the real wage (W/P).

9.3.2 The IS Curve

IS stands for 'Investment-Saving.' The IS curve is defined as the combination of (y, R) that are consistent with equilibrium in the (intertemporal) goods market. In a closed economy, this requires that desired national investment x^D is equal to desired national saving s^D (hence, investment-saving or IS curve).

Chapter 4 discusses at length all the factors that may influence consumer demand (and hence, desired saving). The analysis here simplifies by assuming that desired saving depends primarily on current income y; i.e., $s^D(y)$. For this to make sense, the view must be that any fluctuation in current income is transitory.[2]

Chapter 6 discusses at length all the factors that may influence investment demand. The analysis here utilizes what we learned there and assumes that investment demand depends negatively on the interest rate R and positively on an 'expectation parameter' z^e; i.e., $x^D(R, z^e)$. In Chapter 6, z^e reflected the private sector's forecast of the future return to capital spending. You can think of z^e as shifting for either 'rational' or 'exogenous' reasons (the Keynesian view prefers the latter interpretation).

Equilibrium in the goods market requires:

$$s^D(y) = x^D(R, z^e). \tag{9.3}$$

The IS curve simply represents all the (y, R) combinations that satisfy equation (9.3) for a given z^e. Sometimes, the level of y that satisfies (9.3) is called the *aggregate demand for goods and services* (not to be confused with the AD curve).

- **Exercise 9.2**. Explain why the aggregate demand for goods and services (output) depends negatively on R.

[2]Recall from Chapter 4 that an income change that is perceived to be permanent is not likely to influence desired saving by very much.

From the previous exercise, it follows that y and R are negatively related to each other (for a fixed z^e).

- **Exercise 9.3**. Explain how the aggregate demand for goods and services depends on z^e (explain the economics; do not just describe the mechanics).

9.3.3 The LM Curve

LM stands for 'Liquidity preference - Money supply.' Here, 'liquidity preference' refers to the demand for money. The LM curve is defined as all the (y, R) combinations that are consistent with equilibrium in the money market, for a given supply of real money balances M/P. As such, it bears some resemblance to the AD curve derived earlier, where equilibrium in the money market required $M = Py$. Under this version of the *Quantity Theory*, however, the money market does not depend on R. This is because the simple version of the QTM assumes that the demand for real money balances L depends primarily on real income; i.e., $L(y) = y$.

An intuitive argument can be made, however, that the demand for money should depend on R as well, leading us to write $L(y, R)$. The presumption is that money demand should depend negatively on the nominal rate of interest. In particular, since money earns no interest, the interest rate reflects the opportunity cost of holding money. A higher interest rate is likely to compel individuals to economize on their money holdings (preferring to hold a greater fraction of their wealth in the form of interest-bearing bonds).

Equilibrium in the money market can therefore be expressed by:

$$\frac{M}{P} = L(y, R). \tag{9.4}$$

The LM curve represents all the (y, R) combinations that satisfy equation (9.4), for a given (M/P). Since L is an increasing function of y and a decreasing function of R, it follows that y and R are positively related. That is, since a higher level of income increases money demand, the interest rate must then increase to bring money demand back down to a fixed level (M/P).

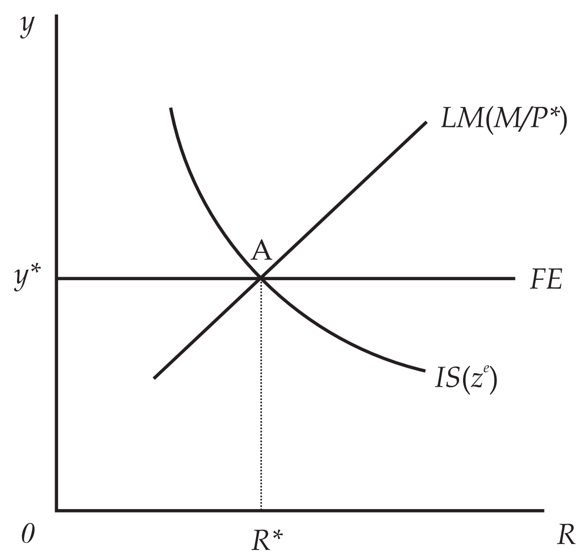
9.3.4 Response to a Money Supply Shock: Neoclassical Model

As before, it is useful to describe the general equilibrium of this model under the assumption that wages and prices are flexible (the neoclassical assumption). The resulting equilibrium can then be interpreted as a 'long-run' scenario.

The general equilibrium of the model consists of a scenario in which the labor, goods, and money market are all in equilibrium simultaneously. Mathematically, the model consists of three equations (IS-LM-FE) and three unknowns (y, R, P).

We want to find a combination of (y, R, P) that satisfy all three equations simultaneously. Such an equilibrium is depicted as point A in Figure 9.3.

FIGURE 9.3
General Equilibrium: Neoclassical Model



In the neoclassical model, an increase in M generates a proportional increase in P (and W), leaving the position of the LM curve unchanged. A money supply shock has no effect on either output or the interest rate; the only effect is raise nominal prices and wages. Money is neutral.

- **Exercise 9.4**. Imagine that the economy receives an 'aggregate demand' shock (i.e., an increase in z^e). Use the logic embedded in Figure 9.3 to argue that such a shock will: (a) leave current output unchanged; (b) increase the interest rate; and (c) increase the price-level. Explain the economics.

- **Exercise 9.5**. Note that the IS-LM-FE analysis is ill-equipped to isolate the 'long-run' effects associated with a current period change in z^e. For example, if z^e has increased this period because individuals are forecasting higher future productivity and if such expectations are correct, higher future productivity will increase the supply of output (shifting the future FE curve 'up'). To the best of your ability, use Figure 9.3 to demonstrate what the future equilibrium may look like.

9.3.5 Response to a Money Supply Shock: Keynesian Model

In the Keynesian model, the economy's 'long-run' general equilibrium position corresponds to the neoclassical case; i.e., point A in Figure 9.3 (i.e., the SRFE curve lies on top of the FE curve). However, if the nominal wage is sticky, money will again be non-neutral in the short-run.

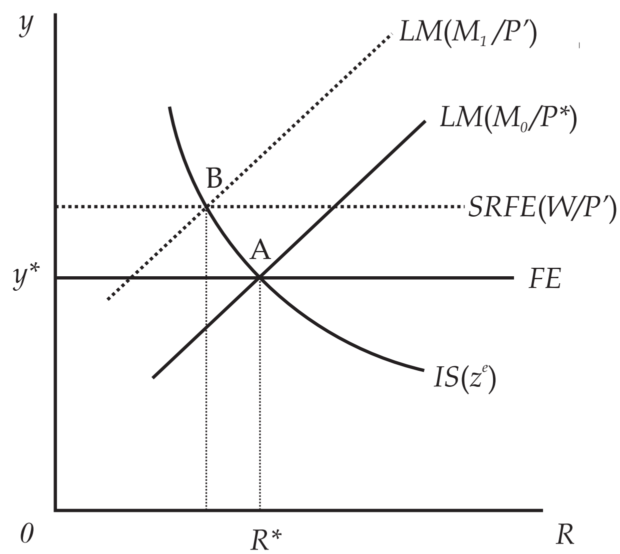
To see how things work here, consider an exogenous increase in the money supply, say from M_0 to M_1. The effect of this shock is to shift the SRFE curve 'up.' That is, for a fixed nominal wage, the resulting increase prices will lower the real cost of labor, hence expanding the supply of output (in the short-run).

Note, however, that the *initial* increase in the price-level is not as large as it would have been in the neoclassical model. This is because the short-run increase in output dampens the price-level response. Since the price-level does not initially rise in proportion to the money supply, the supply of real money balances *increases*, say, from (M_0/P^*) to (M_1/P'), where $P' > P^*$. This increase in real money balances shifts the LM curve 'up,' so that the short-run equilibrium is given by point B in Figure 9.5.

The way a New-Keynesian economist would explain the economics of what is happening here is as follows. The sudden injection of new money leads to a *liquidity effect* that lowers the interest rate (both real and nominal). This liquidity effect occurs because the supply of *real* money balances increases (owing to the partial adjustment in the price-level, brought about by the sticky nominal wage). The lower interest rate then stimulates the aggregate demand for goods and services (the movement up along the IS curve). As the demand for output increases, firms hire more workers to meet the demand (the shift up in the SRFE curve).

In the 'long-run,' (the time it takes workers to renegotiate their nominal wages upward), the nominal wage will rise to its (neoclassical) equilibrium level. This increase in labor costs compels firms to scale back on employment and the supply of output (i.e., the SRFE curve shifts back down to it's original level). The subsequent contraction in output puts further upward pressure on the price-level (shifting the LM curve back down to it's original level). In the long-run, money is neutral (the equilibrium level of output returns to its 'natural' level).

FIGURE 9.4
The Liquidity Effect of a Money Supply Shock



Perhaps one reason why policymakers (central bankers in particular) 'like' the New-Keynesian model is because it implies that monetary policy can influence *real* economic activity–at least, in the short-run. The idea that a central bank might exert such influence is comforting to those who view markets as working 'imperfectly' (sticky nominal variables) and perhaps subject to 'irrational' fluctuations in 'aggregate demand.' According to the New-Keynesian model, the central bank can and should vary its interest rate (money supply) policy to keep the economy as close as possible to its 'natural' level of activity.

- **Exercise 9.6**. Suppose that the economy is initially at its general equilibrium (i.e., point A in Figure 9.4). Imagine further that the economy is subject to an 'aggregate demand' shock (i.e., an increase in z^e).

  (a) Use the logic embedded in Figure 9.4 to work through how the economy may react to such a disturbance, both in the short and long-run.

  (b) Suppose that a central bank interprets the shock as 'irrational exuberance' on the part of the private sector. Explain how the central bank could increase the interest rate to a point that stabilizes GDP at its 'natural' level. Would it be welfare improving for the central bank to act in this manner? Explain.

9.4 How Central Bankers View the World

The IS-LM-FE model developed above captures many of the basic principles that appear to be held by central bankers around the world. The actual theoretical framework employed, however, is an extension of the IS-LM-FE model. The extension involves providing some link between inflation (as opposed to the price-level), inflation expectations, and output. Below, I describe the basic setup of this extended model.

9.4.1 Potential Output

Potential output is defined as that level of output that would be produced in the absence of any nominal rigidities.[3] In a neoclassical model, the economy is always producing 'at potential.' In a New-Keynesian model, the economy may not be producing at potential; e.g., as in point B in Figure 9.4.

Note that potential output is a *theoretical* object; i.e., it is not as if we can simply look at an economy and observe it's potential output. Potential output must be *estimated* within the context of a particular *theory*. The theory adopted by central bankers (largely shared by New-Keynesians) is that potential output grows relatively smoothly over time. Implicit in this view is that the process of technological development occurs in a relatively smooth manner (in contrast to the neoclassical perspective). Furthermore, while various real shocks (like a sudden increase in the world price of oil) may cause shifts in potential output, this source of variability plays a relatively minor role (at least, most of the time) in explaining the business cycle (again, this in contrast to the neoclassical perspective).

Thus, potential output corresponds in some way to an economy's 'trend' level of GDP. The way that one can estimate potential output then is to associate it with the *statistical* trend line running through measurements of the economy's actual GDP growth pattern; for example, as in Figure 9.5.

[3] Potential output is sometimes also referred to as the 'natural' level of output.

FIGURE 9.5
United States 1970.1 - 2003.3

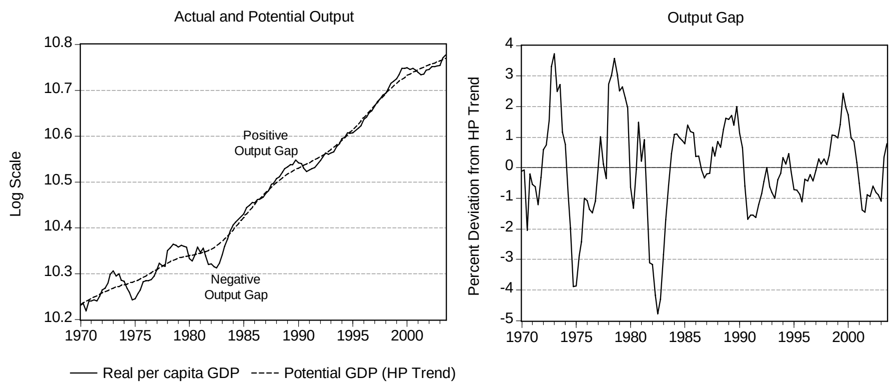


Figure 9.5 suggests that the economy is not usually functioning at potential. The difference between actual and potential output is called the *output gap*. A negative output gap tends to emerge as an economy enters a recession. A positive output tends to emerge after a period of prolonged expansion. Since the (theoretical) output gap can only emerge as a consequence of nominal rigidities, the business cycle here is naturally viewed as being something 'bad;' i.e., the product of a less than perfect market economy. In the 'long-run,' nominal prices adjust to move the economy back to potential. But in the 'short-run,' various shocks to the economy can move the economy away from potential. Since real shocks are presumed to play a relatively minor role, it follows that 'aggregate demand' shocks must constitute the primary source of economic volatility. Hence, to the extent it is possible, policy should endeavor to stabilize the business cycle.[4]

In terms of the theory developed early, you can think of potential (or the 'natural' level) of output as being determined by the neoclassical FE curve. Let y_t^* denote potential output at date t.

[4]It is interesting to contrast the New-Keynesian interpretation of the business cycle with the neoclassical perspective. According to the latter view, economic fluctuations are primarily the product of the natural process of economic development. The market system is viewed as working reasonably well (so that nominal rigidities are not quantitatively important). According to this view then, the economy is always at (or close) to 'potential.' The so-called 'trend' line and 'output gap' identified in Figure 9.5 then is merely the by-product of a *statistical* detrending procedure. One cannot conclude, on the basis of drawing a smooth line through the data, that this smooth line necessarily corresponds to any *theoretical* object. (It is debates like these that make macroeconomics so interesting).

9.4.2 The IS and SRFE Curves

While the economy's potential output is determined by the FE curve, the actual level of output is determined by the IS and SRFE relationships. The way this is usually written is as follows:

$$y_t - y_t^* = -\delta (R_t - R^*) + e_t. \tag{9.5}$$

The term $(y_t - y_t^*)$ represents the output gap. The term $(R_t - R^*)$ denotes an 'interest rate gap.' In this latter expression, you can think of R^* as being determined by 'long-run' neoclassical considerations; so that R^* can be thought of as the 'natural' rate of interest. As is the case with output, the actual (real) interest rate may differ from its natural rate. The relationship in (9.5) asserts that the output gap is negatively related to the interest rate gap, where $\delta > 0$ is just a parameter that indexes the strength of this relationship. The equation (9.5) actually combines the IS and SRFE relationships together. It has the flavor of an IS curve because equilibrium in the goods market requires that (y_t, R_t) be negatively related to each other (i.e., an increase in the interest rate lowers the aggregate demand for goods and services). But since y_t has the interpretation of being the actual level of output (as opposed to just the theoretical demand for output), implicit in (9.5) is the idea that firms actually deliver the output demanded (i.e., the economy is on the SRFE or SRAS curve).

The term e_t in (9.5) is meant to represent an 'aggregate demand shock.' One interpretation of this shock is that it reflects 'animal spirits' that lead to exogenous fluctuations in investment demand. Alternatively, one can think of e_t as representing an exogenous government spending shock.

9.4.3 The Phillips Curve

The Phillips curve provides a link between inflation and output. The way this is usually written is as follows:

$$\Pi_t = \Pi_t^e + \phi (y_t - y_t^*) + u_t. \tag{9.6}$$

Equation (9.6) asserts that the inflation rate is determined in part by expected inflation, the output gap, and possibly other forces ('inflation shocks').

In the IS-LM-FE model studied above, a positive 'aggregate demand shock' (e.g., an increase in z^e) tends to put upward pressure on the price-level. This same shock is also likely to drive actual output above potential; see Exercise 9.6. This type of reasoning suggests that there is a positive relationship between inflation and the output gap; i.e., $\phi > 0$.

The idea that expectations of inflation may influence the actual inflation rate is something we touched on briefly in Section 8.3.1. (the discussion there was in terms of the price-level). In particular, if the money supply is endogenous in the sense of being determined entirely by the demand for money at a given interest rate (as will be the case in this model), then there is an element of nominal indeterminacy present. In particular, an expected inflation can become a self-fulfilling prophesy.[5]

Precisely how inflation expectations are modeled appears to vary across different New-Keynesian models. One view holds that expectations are 'backward-looking' in the sense that people form their expectations of inflation primarily on a recent history of actual inflation rates. Another view holds that expectations are 'forward-looking' in the sense of depending on (among other things) the entire future path monetary policy actions. Some New-Keynesian models embody both backward and forward looking elements. For our purpose, the simplest thing to assume is backward-looking behavior; for example:

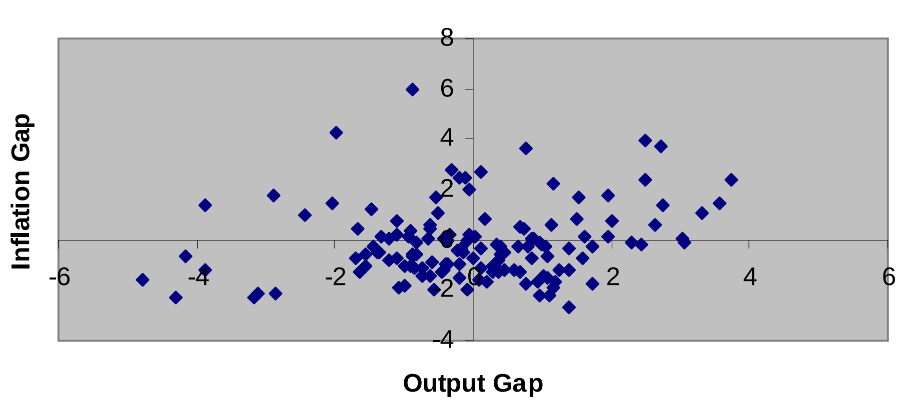
$$\Pi_t^e = \Pi_{t-1}. \tag{9.7}$$

In this case, the Phillips curve relationship in (9.6) may be rewritten as follows:

$$\Pi_t = \Pi_{t-1} + \phi\left(y_t - y_t^*\right) + u_t. \tag{9.8}$$

Figure 9.6 plots the 'inflation gap' $(\Pi_t - \Pi_t^e)$ against the output gap $(y_t - y_t^*)$ for the U.S. economy (1970.1–2003.3).[6] Empirically, the Phillips curve relationship does not appear to be as tight as one might like, although there is a positive correlation (0.196) between the gap measures.

Figure 9.6: Phillips Curve



[5] In some New-Keynesian models, expected inflation influences actual inflation through a strategic complementarity that is present in the pricing decisions of monopolistically competitive firms.

[6] In Figure 9.6, I use the Livingston survey of inflation expectations for my measure of Π_t^e.

As for the 'inflation shock' term u_t, I am not very clear as to what this is supposed to represent in reality.[7] Perhaps one way to think of this shock is as influencing the formation of inflation expectations; i.e., $\Pi_t^e = \Pi_{t-1} + u_t$. Such an interpretation is not entirely inconsistent with the perception of some central bankers that inflation expectations have partly a life of their own (i.e., are driven in part by animal spirits).

9.4.4 Monetary Policy: The Taylor Rule

Central bankers view themselves not as determining the money supply, but as determining 'the' nominal interest rate R_t^n. Given this choice of interest rate, the money supply is determined endogenously by the private sector's demand for money. In effect, the LM curve becomes vertical in Figure 9.4).

Of course, Figure 9.4 depicts the real interest rate on the x-axis. A change in the nominal interest rate will only affect the real interest rate to the extent that inflation expectations do not move one-for-one with changes in the nominal rate. Given our assumption concerning inflation expectations in (9.7), this will in fact be the case. In the short-run then, the nominal and real interest rates move together; i.e., from the Fisher equation, $R_t = R_t^n / \Pi_t^e$.

The central bank is imagined to have a long-run inflation target Π^*, which serves as a 'nominal anchor' (in the absence of such a target, the equilibrium inflation rate would be indeterminate).

Now, imagine that there were no shocks in this economy; i.e., $e_t = u_t = 0$, for all t (with y_t evolving in a smooth and deterministic manner). In the absence of any shocks, it makes sense to think of inflation expectations as being determined by the inflation target; i.e., $\Pi^e = \Pi^*$. Then, given the inflation target Π^*, the best a central bank can do is to set the nominal interest rate to a level that would prevail in the neoclassical model; i.e., $R^n = \Pi^* R^*$. In such a world, the level of output would always be at potential.

- **Exercise 9.7**. What would happen in such a world if the central bank lowered the nominal interest rate below $\Pi^* R^*$ (say, for one period)?
- **Exercise 9.8**. Suppose that the central bank wished to lower its inflation target. Explain how the central bank could achieve this goal, but at the cost of a recession.

Now, imagine that the economy is inflicted with 'aggregate demand shocks' e_t. Consider, for example, an exogenous increase in aggregate demand. For a given interest rate (both real and nominal), the effect of this shock is expand

[7] These shocks are sometimes referred to as 'mark-up shocks,' which influence the pricing-decisions of monopolistically competitive firms. I would appreciate someone providing me some good real-world examples of a 'mark-up' shock as well as evidence suggesting their importance for driving business cycle behavior.

output and the price level (since the nominal wage is rigid). The expansion in output above potential is welfare-reducing. Furthermore, the surprise increase in the price-level is inflationary (i.e., since actual inflation is higher than expected, individuals subsequently update their inflation expectations to a higher level–which generates a permanently higher inflation rate). To bring the inflation rate back down to the target level, the central bank would have to generate a recession; i.e., re: Exercise 9.8. The way in which a central bank may prevent all of this from happening is to *respond* to the aggregate demand shock by temporarily increasing the interest rate. Such a response would 'frustrate' the 'excess demand' and 'inflationary pressure' brought on by the aggregate demand shock.[8]

If aggregate demand shocks were the only shocks hitting an economy, then the central bank's job would be relatively easy. That is, it would simply increase the interest rate when it observed output increasing beyond potential, and decrease the interest rate when output it observed output moving below potential. Such a policy would stabilize both output and inflation.

But if there are other shocks hitting the economy, then the central bank's job is not so straightforward. Consider, for example, the arrival of a positive 'inflation shock' u_t. Left unchecked, the effect of such a shock would be to increase the inflation rate permanently (via the adjustment in expectations). From the Phillips curve relation (9.8), we see that the only way to keep the inflation rate stable when $u_t > 0$ is to make $(y_t - y_t^*) < 0$. In other words, to stem the inflationary pressure of an inflation shock, the central bank must raise the interest rate, thereby generating a recession (to 'cool off' aggregate demand). Any long-term benefit from maintaining the inflation target at Π^* must therefore be weighed against the short-run cost of making output move away from potential.

In sum, we see that when both types of shocks are operating, the central bank faces a trade-off between stabilizing output (around potential) and stabilizing the inflation rate (around the inflation target). A central bank that is more concerned with stabilizing output will be less concerned about controlling inflation. In a sense, such policy places more weight on the 'short-run' relative to the 'long-run.' Conversely, a central bank that is more concerned with stabilizing the inflation rate will be less concerned about stabilizing output around potential. In a sense, such a policy places more weight on the 'long-run' relative to the 'short-run.'

Let $0 \leq \lambda \leq 1$ denote a parameter that indexes the degree to which a central bank is concerned about stabilizing inflation. Then an 'optimal' interest rate policy can be thought of taking the form:

$$R_t^n = \Pi^* R^* + (1 - \lambda)(y_t - y_t^*) + \lambda (\Pi_t - \Pi^*). \tag{9.9}$$

This equation is an example of a *Taylor rule* (named after John Taylor, who first

[8] In terms of Figure 9.4, the central bank would in effect be keeping the economy along the FE curve.

proposed the equation as a description of how U.S monetary policy appeared to be conducted; see Taylor, 1993). The Taylor rule suggests that a central bank should increase the interest rate if as output rises above potential and/or as inflation rises above the inflation target. How strongly a central bank should react to either the output gap or the inflation gap will depend on the preferences of the central bank as indexed by the parameter λ.

9.5 Summary

The New-Keynesian view accepts the neoclassical model as a reasonable way to think about how a macroeconomy functions in the 'long-run.' However, New-Keynesians argue that in the 'short-run,' the economy may function quite apart from neoclassical considerations. The primary reason for this is because markets do not work perfectly well in the short-run. In particular, there are 'contracting frictions' that result in nominal rigidities, leaving the price-mechanism unable to cope with allocating resources efficiently at every instant in response to exogenous disturbances. Furthermore, these same nominal rigidities imply that monetary policy can (and perhaps should) influence real activity.

The New-Keynesian view also differs from the neoclassical view in terms of the emphasis placed on which type of 'shock' is primarily responsible for the business cycle. The neoclassical view emphasizes productivity shocks (and other real disturbances). Such shocks have both 'supply' and 'demand' effects; the latter which occur via 'rational' fluctuations in expectations. In contrast, the New-Keynesian view downplays the role of productivity shocks and emphasizes the role played by exogenous 'demand' shocks. Since these shocks are not viewed as being correlated with future productivity, they are likely interpreted as being generated largely by 'animal spirits.'

In terms of monetary policy, an extreme neoclassical view might suggest the abandonment of a central bank in favor of a 'free-banking' system. Since this is unlikely to be a practical alternative in the foreseeable future, most neoclassical economists would likely favor a monetary policy that targets the inflation rate at some 'low' level (possibly negative). In contrast, New-Keynesian economists are likely to favor an interest rate policy that helps stabilize output at least to some extent, while making sure that inflation does not get wildly out of control.

Which of these two views serve as a better approximation to reality is still a matter of debate (at least, in academia). But there is little doubt that the New-Keynesian perspective is more widely held, especially among policymakers. Nevertheless, it is ironic that despite these differences in perspective, many central banks have recently veered the conduct of their monetary policy strongly in favor of inflation-targeting–i.e., a policy that most neoclassical economists would likely favor themselves.

9.6 References

1. Bills, Mark and Peter Klenow (2002). "Some Evidence on the Importance of Sticky Prices," NBER working paper #9069.

2. Caplin, Andrew and Daniel Spulber (1987). "Menu Costs and the Neutrality of Money," *Quarterly Journal of Economics*, 102(4): 703–726.

3. Klenow, Peter and O. Krystov (2003). "State-Dependent or Time-Dependent Pricing: Does it Matter for Recent U.S. Inflation?" NBER working paper #11043.

4. Taylor, John B. (1993). "Discretion Versus Policy Rules in Practice," *Carnegie-Rochester Conference Series on Public Policy*, 39, pp. 195-214.

9.A Are Nominal Prices/Wages Sticky?

At one level, the answer to this question seems obvious: of course they are. Most people likely have in mind the behavior of their own wage when they answer in this way. Nominal wage rates can often remain fixed for several months on end in some professions. Likewise, the prices of many products appear not to change on a daily or even monthly basis (e.g., newspaper prices, taxi fares, restaurant meals, etc.). As is so often the case, however, first impressions are not always correct; and, if they are, they do not always lead to an obvious conclusion.

Let us first consider the evidence on product prices, which is based on the empirical work of Bils and Klenow (2002) and Klenow and Krystov (2003):

TABLE 9.1
Duration of Prices by Category of Consumption, 1995–97

Category	Median Duration (months)	Share of CPI (percent)
All Items	4.3	71.2
Goods	3.2	30.4
Services	7.8	40.8
Food	3.4	17.1
Home Furnishing	3.5	14.9
Apparel	2.8	5.3
Transportation	1.9	15.4
Medical Care	14.9	6.2
Entertainment	10.2	3.6
Other	6.4	7.2

According to Table 9.1, the median duration of price 'stickiness' is about 4.3 months across a broad range of product categories. Thus, it does appear to be the case that *individual* product prices display a form of nominal stickiness.

But one should be careful here. Price-stickiness at the *individual* level need not translate into price-level stickiness (i.e., price-stickiness at the *aggregate* level). Suppose, for example, that all firms keep their prices fixed for 4 months, but when they do change their prices, they change them significantly. One way for the price-level to display stickiness is if a large number of firms *synchronized* their price changes (for example, if all firms changed their prices at the same time–once every 4 months). On the other hand, if firms changed their prices in a completely *unsychronized* way (e.g., if everyday a small number of firms change their prices), then the price-level may actually be flexible, despite any inflexibility at the individual level.[9]

[9] This result also requires that firms adopt an optimal state-contigent pricing-rule; see

It is probably fair to say, however, that reality lies somewhere between these two extreme cases. If it does, then two other questions immediately present themselves. The first question is *where* in between these two extremes? Are we talking two, three, or possibly four months? Suppose that it is three months (i.e., one quarter). If the price-level is sticky for one quarter, then the second question is whether this 'long enough' to have any important and lasting macroeconomic implications? As things stand, the jury is still out on this question.

If price-level stickiness is an important feature of the economy (and it may very well be), one is left to wonder about the source of price stickiness.[10] One popular class of theories postulates the existence of 'menu costs,' that make small and frequent price changes suboptimal behavior.[11] The idea behind a fixed cost associated with changing prices seems plausible enough. But the theory is not without its problems. in particular, Table 9.1 reveals a great deal of variation in the degree of price-stickiness across product categories. Casual empiricism suggests that this is the case. For example, you may have noticed that the price of gasoline at your local gas station fluctuates on almost a daily basis. At the same time, this same gas retailer keeps the price of motor oil fixed for extended periods of time. Why is it so easy to change the price of gasoline but not motor oil? Does motor oil have a larger menu cost?

Much of what I have said above applies to nominal wages as well. While some nominal wages appear to be sticky (e.g., my university salary is adjusted once a year), others appear to be quite flexible. For example, non-union construction workers, who often charge piece rates that adjust quickly to local demand conditions). Furthermore, from Chapter 7 we learned that there are huge flows of workers into and out of employment each month (roughly 5% of the employment stock). It is hard to imagine that negotiated wages remain 'inflexible' to macroeconomic conditions at the time new employment relationships are being formed. Given the large number of new relationships that are being formed each month, it is even more difficult to imagine how the average nominal wage can remain 'sticky' for any relevant length of time. These challenges to the sticky price/wage hypothesis are the subject of ongoing research.

Caplin and Spulber (1987).

[10] Understanding the forces that give rise to price stickiness is important for designing an appropriate policy response.

[11] A menu cost refers to a fixed cost; for example, the cost of printing a new menu everyday with only very small price differences.

Chapter 10

The Demand for Fiat Money

10.1 Introduction

Earlier, we defined money to be any object that circulated widely as a means of payment. We also noted that the vast bulk of an economy's money supply is created by the private sector (i.e., chartered banks and other intermediaries). The demand-deposit liabilities created by chartered banks are debt-instruments that are, like most private debt instruments, ultimately backed by real assets (e.g., land, capital, etc.). If a chartered bank should fail, for example, your demand-deposit money represents a claim against the bank's assets.

In most modern economies, a smaller, but still important component of the money supply constitutes small-denomination government paper notes called *fiat money*. Fiat money is defined as an intrinsically useless monetary instrument that can be produced at (virtually) zero cost and is unbacked by any real asset.

Now, let's stop and think about this. If fiat money is intrinsically useless, what gives it value? In other words, why is the demand for (fiat) money not equal to zero? You can't consume it (unlike some commodity monies). It doesn't represent a legal claim against anything of intrinsic value (unlike private monetary instruments). Furthermore, a government can potentially print an infinite supply of fiat money at virtually zero cost of production (there are many historical examples). Explaining why fiat money has value is not as straightforward as one might initially imagine.

Both the *Quantity Theory* and the New-Keynesian model studied above simply *assume* that fiat money has value (i.e., they simply assume that the demand for fiat money is positive). In this chapter, we develop a theory that shows under what circumstances fiat money might have value (without assuming the result). This theory is then applied to several interesting macroeconomic issues.

10.2 A Simple OLG Model

The basic setup here was first formulated by Samuelson (1958) in his now famous *Overlapping Generations* (OLG) model. Consider a world with an infinite number of time-periods, indexed by $t = 1, 2, 3, ..., \infty$. The economy consists of different types of individuals indexed by $j = 0, 1, 2, ...\infty$. In Samuelson's original model, these different types are associated with different 'generations' of individuals. Each generation (with the exception of the initial generation) was viewed as living for two periods. At each point in time then, the economy was viewed as having a set of 'young' and 'old' individuals, who may potentially want to trade with each other. As will be shown, however, one need not interpret different types literally as 'generations' (although, the original labelling turns out to be convenient).

Consider a member of generation $j = t$, for $j > 0$. This person is assumed to have preferences for two time-dated goods $(c_1(t), c_2(t+1))$, which we can represent with a utility function $u(c_1(t), c_2(t+1))$. You can think of c_1 as representing consumption when 'young' and c_2 as representing consumption when 'old.' There is also an 'initial old' generation ($j = 0$) that 'lives' for only one period (at $t = 1$); this generation has preferences only for $c_2(1)$.

Each generation has a non-storable endowment $(y_1(t), 0)$. That is, individuals are endowed with output when young, but have no output when old (the initial old have no endowment). Thus, the intergenerational pattern of preferences and endowments can be represented as follows:

	Time					
Generation	1	2	3	4	5	$\rightarrow$
0	0					
1	$y_1(1)$	0				
2		$y_1(2)$	0			
3			$y_1(3)$	0		
$\downarrow$				$y_1(4)$	0	

Note that instead labelling a type j individual as belonging to a 'generation,' one can alternatively view all individuals as living at the same time but with specialized preferences and endowments. For example, a type $j = 1$ individual is endowed with a time-dated good $y_1(1)$. This individual also has preferences for two time-dated goods: $y_1(1)$ and $y_1(2)$. Clearly, this individual values his own endowment. But he also values a good which he does not have; i.e., $y_1(2)$. This latter good is in the hands of individual $j = 2$. Note that whether individuals lives forever or for only two periods does not matter here. What matters is that at any given date, there is a complete lack of double coincidence of wants. In particular, note that individual $j = 2$ does not value anything that individual $j = 1$ has to offer. This lack of double coincidence holds true at every date (and would continue to hold true if individuals lived arbitrarily long lives).

10.2.1 Pareto Optimal Allocation

As in the Wicksellian model studied earlier, this economy features a lack of double coincidence of wants. For example, the initial old generation wants to eat, but has no endowment. The initial young generation has output which the initial old values, but the initial old have nothing to offer in exchange. Likewise, the initial young would like to consume something when they are old. Since output is nonstorable, they can only acquire such output from the second generation. However, the initial young have nothing to offer the second generation of young. In the absence of any trade, each generation must simply consume their endowment; this allocation is called *autarky*.

However, as with the Wicksellian model, this lack of double coincidence does not imply a lack of gains from trade in a collective sense. To see this, let us imagine that all individuals (from all 'generations') could get together and agree to cooperate. Equivalently, we can think of what sort of allocation a social planner might choose. In this cooperative, the initial young would make some transfer to the initial old. Thus, the initial old are made better off. But what do the initial young get in return for this sacrifice? What they get in return is a similar transfer when they are old from the new generation of young. If this pattern of exchanges is repeated over time, then each generation is able to smooth their consumption by making the appropriate 'gift' to the old. Let me now formalize this idea.

Let N_t denote the number of individuals belonging to generation $j = t$. At any given date t then, we have N_t 'young' individuals and N_{t-1} 'old' individuals who are in a position to make some sort of exchange. The total population of traders at date t is therefore given by $N_t + N_{t-1}$. We can let this population grow (or shrink) at some exogenous rate n, so that $N_t = nN_{t-1}$. If $n = 1$, then the population of traders remains constant over time at $2N$. For simplicity, assume that the endowment of goods is the same across generations; i.e., $y_1(t) = y$.

Assume that the planner wishes to treat all generations in a symmetric (or 'fair') manner. In the present context, this means choosing a consumption allocation that does not discriminate on the basis of which generation an individual belongs to; i.e., $(c_1(t), c_2(t+1)) = (c_1, c_2)$. In other words, in a symmetric allocation, every person will consume c_1 when young and c_2 when old (including the initial old). Thus the planner chooses a consumption allocation to maximize $u(c_1, c_2)$.

At each date t, the planner is constrained to make the consumption allocation across young and old individuals by the amount of available resources. Available output at date t is given by $N_t y$. Consequently, the resource constraint is given by:

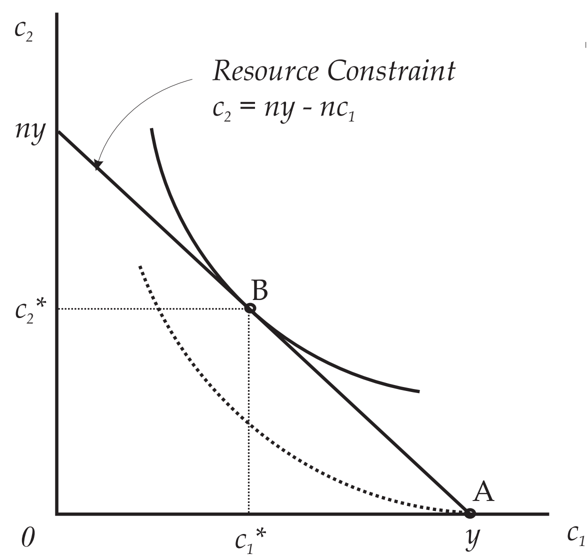
$$N_t c_1 + N_{t-1} c_2 = N_t y;$$

or, by dividing through by N_t :

$$c_1 + \frac{c_2}{n} = y. \qquad (10.1)$$

Notice that this resource constraint looks very much like the lifetime budget constraint studied in earlier chapters, with n playing the role of the real interest rate. The solution to the planner's problem is depicted as point B in Figure 10.1.

FIGURE 10.1
Pareto Optimal Allocation in an OLG Model



Point A in Figure 10.1 depicts the autarkic (no trade) allocation. Clearly, all individuals can be made better off by partaking in a system of intergenerational trades (as suggested by the planner). The initial young are called upon to sacrifice $(y - c_1^*)$ of current consumption, which is transferred to the initial old. This sacrifice is analogous to an act of saving, except that no private debt contract (between creditor and debtor) is involved. In particular, note that the initial old will never pay back their 'debt.' This sacrificial act on the part of the initial young is repaid not by the initial old, but by the next generation of young; and so on down the line.

One way to imagine how trade takes place is to suppose that a central planner forcibly takes (i.e., taxes) the initial young by the amount $(y - c_1^*)$ and transfers (i.e., subsidizes) the initial old by this amount. This pattern of taxes and subsidies is then repeated at every date. Under this interpretation, the planner is behaving as a government that is operating a *pay-as-you-go* social security system.

But there is another way to imagine how such exchanges may take place *voluntarily*. What is required for this is the existence of a *centralized public record-keeping system*. In particular, imagine that the young of each generation adopt a strategy that involves making transfers to those agents who *have a record of having made similar transfers in the past*. The availability of a public record-keeping system makes individual trading histories accessible to all agents. Under this scenario, the initial young would *willingly* make a transfer to the initial old. Why is this? If they do, their sacrifice is recorded in a public data bank so that they can be identified and rewarded by future generations. If they do not make the sacrifice, then this too is recorded but is in this case punished by future generations (who will withhold their transfer). By not making the sacrifice, an individual would have to consume their autarkic allocation, leaving them worse off.

The key thing to note here is that an optimal private trading arrangement can exist despite the lack of double coincidence of wants and without anything that resembles money. As Kocherlakota (1998) has stressed, money is not necessary in a world with perfect public record-keeping.

10.2.2 Monetary Equilibrium

Imagine now that there is no planner and that there is no public record-keeping technology. Since individual sacrifices cannot be recorded (and hence rewarded), the only (non-monetary) equilibrium in this model is autarky. In such a world, however, there is now a potential role for fiat money. Predictably, the role of fiat money is to substitute for the missing public recording-keeping technology.

To describe a monetary equilibrium, we proceed in two steps. First, we will describe individual decision-making under the *conjecture* that money has value. Second, we will verify that such a conjecture can be consistent with a rational expectations equilibrium. In what follows, we will restrict attention to stationary equilibria; i.e., allocations such that $(c_1(t), c_2(t+1)) = (c_1, c_2)$.

Imagine that the initial old are endowed with M units of fiat money (perhaps created and distributed by a government). The supply of fiat money is assumed to be fixed over time. We also assume that fiat money cannot be counterfeited. The basic idea here is to get the initial young to sell some of their output to the initial old in exchange for fiat money. This will turn out to be rational for the initial young if they expect fiat money to have value in the future.

Let p_t denote the price-level at date t (i.e., the amount of money it takes to purchase one unit of output). Assume that individuals view the time-path of the price-level as exogenous and conjecture that money is valued at each date (so that $p_t < \infty$). Later we will have to verify that such a conjecture is rational. But for now, given prices, a young agent faces the following budget constraint:

$$p_t c_1 + m_t = p_t y;$$

where m_t denotes 'saving' in the form of fiat money. Thus, given a price-level p_t, a young individual has nominal income $p_t y$. Some of this income can be used to purchase consumption $p_t c_1$ and the remainder can be used to purchase money m_t (from the old). Dividing through by p_t, this equation can alternatively be written as:

$$c_1 + q = y; \tag{10.2}$$

where $q \equiv m_t / p_t$ represents an individual's real money balances (the current purchasing power of the money they acquire).

Note that while output is nonstorable, money can be carried into the future. Thus, in the second period of life, an individual faces the following budget constraint:

$$\begin{aligned} p_{t+1} c_2 &= m_t; \qquad (10.3)\\ c_2 &= \left(\frac{p_t}{p_{t+1}} \right) \left(\frac{m_t}{p_t} \right); \\ c_2 &= \Pi^{-1} q; \end{aligned}$$

where $\Pi \equiv p_{t+1}/p_t$ denotes the (gross) inflation rate.

- **Exercise 10.1**. Let $v_t = 1/p_t$ denote the value of money (i.e., the amount of output that can be purchased with one unit of money). Let $R = (v_{t+1}/v_t)$ denote the (gross) real rate of return on money. Show that the real return on money is inversely related to the inflation rate Π.

By combining (10.2) and (10.3), we can derive a young individual's lifetime budget constraint:

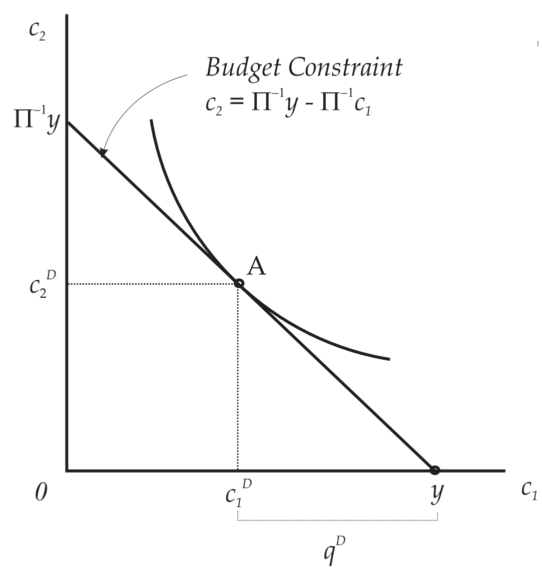
$$c_1 + \Pi c_2 = y. \tag{10.4}$$

This expression should look familiar to you; i.e., see Chapter 4. In particular, by substituting $\Pi = 1/R$ one derives $c_1 + R^{-1} c_2 = y$. The only difference here is that R does not represent an interest rate on a private security; i.e., it represents the rate of return on fiat money (i.e., the inverse of the inflation rate).

Now take a closer look at equation (10.4). Notice that the inflation rate looks like the 'price' of c_2 measured in units of c_1. And, indeed it is. Think of c_1 as representing a 'non-cash' good (i.e., a good that can be purchased without money) and think of c_2 as representing a 'cash' good (i.e., a good that can only be purchased by first acquiring money). We see then that a high inflation rate corresponds to a high price for the cash good (relative to the non-cash good). In particular, if the inflation rate is infinite, then the price of acquiring the cash good is infinite (it makes no sense to acquire cash today since it will have zero purchasing power in the future).

Given some inflation rate Π, a young person seeks to maximize $u(c_1, c_2)$ subject to the budget constraint (10.4). The solution to this problem is a pair of demand functions $c_1^D(y, \Pi)$ and $c_2^D(y, \Pi)$. The demand for real money balances is then given by $q^D(y, \Pi) = y - c_1^D(y, \Pi)$. This solution is depicted as point A in Figure 10.2.

FIGURE 10.2
Money Demand in an OLG Model



- **Exercise 10.2**. Show that the demand for real money balances may be either an increasing or decreasing function of the inflation rate. Explain. However, make a case that for very high rates of inflation, the demand for money is likely to go to zero (in particular, show what happens as Π goes to infinity).

What we have demonstrated so far is that if $\Pi < \infty$, then there is a positive demand for fiat money. However, since we have not explained where Π comes from, the theory is incomplete. In particular, we do not know at this stage whether $\Pi < \infty$ is consistent with a rational expectations equilibrium.

In a competitive rational expectations equilibrium, we require the following:

1. Given some *expected* inflation rate Π, individuals choose $q^D(y, \Pi)$ optimally;

2. Given the behavior of individuals, markets clear at every date; and
3. The *actual* inflation rate Π is consistent with expectations.

We have already demonstrated what is required for condition (1). Condition (2) argues that at each date, the supply of money must be equal to the demand for money. Mathematically, we can write this condition as:

$$\frac{M}{p_t} = N_t q^D(y, \Pi). \tag{10.5}$$

Since this condition must hold at every date, the following must also be true:

$$\frac{M}{p_{t+1}} = N_{t+1} q^D(y, \Pi).$$

Dividing the former by the latter, we derive:

$$\frac{p_{t+1}}{p_t}\frac{M}{M} = \frac{N_t}{N_{t+1}}\frac{q^D(y, \Pi)}{q^D(y, \Pi)};$$

which simplifies to:

$$\Pi^* = \frac{1}{n}. \tag{10.6}$$

What this tells us is that *if* individuals expect an inflation rate $\Pi^* = n^{-1}$, then there is a competitive rational expectations equilibrium in which the actual inflation rate turns out to be Π^*. Note that in this case, the *equilibrium budget line* in Figure 10.2 corresponds precisely to the resource constraint in Figure 10.1. In other words, the resulting equilibrium is Pareto optimal.

However, this is not quite the end of the story. As it turns out, this is not the only rational expectations equilibrium in this model. The previous equilibrium was constructed under the assumption that individuals initially expected that money would retain some positive value over time; i.e., that $p_t < \infty$ for all t. Imagine, on the other hand, that individuals initially believe that fiat money retains no future purchasing power; i.e., that $p_{t+1} = \infty$. In this case, a rational individual would not choose to sell valuable output for money that he expects to be worthless; i.e., $q^D = 0$. But from the money market clearing condition (10.5), if $q^D = 0$ then $p_t = \infty$.

In other words, if everyone believes that fiat money will have no value, then in equilibrium, this belief will become true (and hence, is consistent with a rational expectations equilibrium). Fiat money can only have value if everyone believes that it will. This is another example of a *self-fulfilling prophesy* phenomenon that was discussed in a different context in Chapter 2 (see Section 2.6.2).

- **Exercise 10.3**. Explain why the value of a fully-backed monetary instrument is not likely to depend on a self-fulfilling prophesy. Hint: try to see whether multiple rational expectations equilibria are possible in the Wicksellian model studied in Chapter 8.

10.3 Government Spending and Monetary Finance

In Chapter 5, we examined how a government could finance a given expenditure stream using either taxes or debt (the promise of future taxes). But for a government with control over the printing press (the supply of small denomination paper notes), there is a third way in which it may finance its spending needs: through the creation of new money. In this section, we study the economic implications of financing government spending requirements through money creation.

According to the *Quantity Theory of Money*, any expansion in the money supply is inflationary. The evidence supporting this prediction, however, appears to be mixed; i.e., see Smith (1985). Perhaps one reason for this is because historically, governments typically backed their money with gold (a so-called 'gold-standard' regime). Money, therefore, resembled a government bond, since it reflected a promise to deliver gold at some time in the future (presumably via higher future taxes).

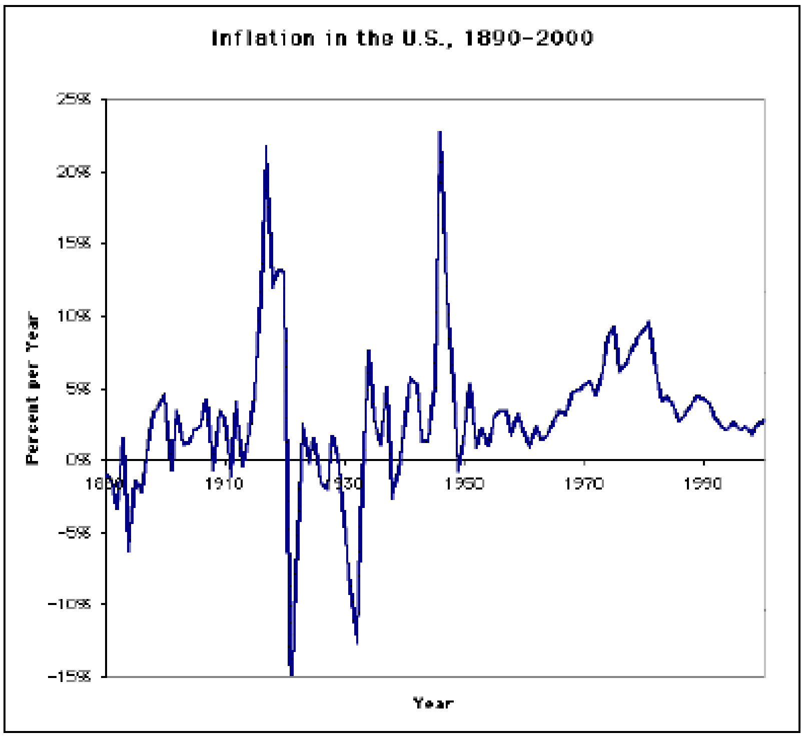
During periods of fiscal crisis (often during times of war), governments would temporarily abandon the redemption of government money for gold, promising to restore redemption some time in the future. During these episodes, the value of money would typically fall (i.e., the price-level would rise), and would otherwise behave much like the value of a bond in partial default. The value of money in these episodes appeared to depend on the credibility of the government's promise to restore the future redemption of its currency. In some cases, such as the Confederate government during the U.S. civil war (1861-65), the government's promise of future redemption was ultimately reneged on (so that Confederate notes lost all of their value by the end of the war). In other cases, governments made good their promise of redemption (at least partially), which often led to a rise in the value of money (a fall in the price-level).

Figure 10.3 plots the behavior of the inflation rate in the United States from 1890–2000.[1] In the early part of the sample, the U.S. was on a gold standard. This gold standard was temporarily abandoned when the U.S. entered the first world war in 1917 (along with many of the war's major belligerents). Notice the large spike in inflation during and just following the war. The gold standard was resumed for a brief period during 1925-31. Notice how the 1920s and early 1930s were characterized by deflation (with the 1920s being a period of prosperity–the so-called 'roaring 20s'). During the depths of the great depression, the U.S. again abandoned the gold standard. In 1941, the United States entered into the second world war. Price-controls kept the inflation rate artificially low during this period. Once the price controls were lifted (at the end of the war), inflation spiked again. Following the second world war, the U.S. again adopted a gold standard (via the Bretton-Woods fixed exchange rate system). Under this gold standard, most countries settled their international balances in U.S. dollars, but

[1] Source: www.j-bradford-delong.net/ Econ_Articles/ woodstock/ woodstock4.html

the U.S. government promised to redeem other central banks' holdings of dollars for gold at a fixed rate of \$35 per ounce. However, persistent U.S. balance-of-payments deficits (owing largely to the fiscal pressures brought on by the U.S. war in Vietnam) steadily reduced U.S. gold reserves, reducing confidence in the ability of the United States to redeem its currency in gold. Finally, on August 15, 1971, President Nixon announced that the United States would no longer redeem currency for gold. Inflation remained persistently high throughout the 1970s, until it was finally brought under control in the early 1980s.

FIGURE 10.3



Thus, the figure above suggests that periods of extraordinarily high inflation are linked to periods of fiscal crisis. During a fiscal crisis, the government needs to acquire resources. If political pressures limit the government's ability to acquire resources through direct taxes or conventional bond issues, it may instead resort to printing small denomination paper notes. If these paper notes are viewed as largely fiat in nature, then the extra supply of money is likely to depress the value of money (in accordance with the Quantity Theory). The resources that the government acquires in this manner is called *seigniorage revenue* or the *inflation tax*.

Nowadays, most governments issue only fiat money. In most developed economies, the inflation rate (and hence, the inflation tax) is relatively low (at least, during peacetime). In many other countries, however, the ability to tax or issue bonds is severely limited, so that the inflation tax constitutes a more important source of government revenue. One question that immediately springs to mind is whether the monopoly control of fiat money gives a government an unlimited ability to raise seigniorage revenue. This question can be investigated in the context of the model developed above.

10.3.1 The Inflation Tax and the Limit to Seigniorage

Imagine that there is a government that requires $G_t = gN_t$ units of output in period t, where $0 < g < y$ represents the amount of government spending per young person. To focus on money creation as a revenue device, let us abstract from taxes and bonds. The only way in which the government can pay for its purchases is by printing new money: $M_t - M_{t-1}$. Let μ denote the gross rate of money growth; i.e., $M_t = \mu M_{t-1}$. Thus, the amount of new money printed in period t can be written as:

$$M_t - M_{t-1} = M_t - \frac{M_t}{\mu} = \left(1 - \frac{1}{\mu}\right) M_t.$$

The government's budget constraint in every period can therefore be written as:

$$p_t G_t = \left(1 - \frac{1}{\mu}\right) M_t.$$

Since $G_t = gN_t$, this can further be simplified as follows:

$$g = \left[1 - \frac{1}{\mu}\right] \frac{M_t}{N_t p_t}. \tag{10.7}$$

Now, for a given inflation rate Π, optimal behavior on the part of households implies a money demand function $q^D(y, \Pi)$; i.e., see Figure 10.2. From the money market clearing condition, we know that the following must be true at every date:

$$\frac{M_t}{p_t} = N_t q^D(y, \Pi); \tag{10.8}$$
$$\frac{M_{t+1}}{p_{t+1}} = N_{t+1} q^D(y, \Pi).$$

Dividing the former equation by the latter and rearranging terms, we derive an expression for the equilibrium inflation rate:

$$\Pi^* = \frac{\mu}{n}. \tag{10.9}$$

Notice that this expression corresponds to equation (10.6) for the case of a constant money supply; i.e., $\mu = 1$. Note further that this expression is very similar to the one implied by the *Quantity Theory of Money*; i.e., see Chapter 8 (section 8.4).

We can now combine equations (10.7), (10.8) and (10.9) to form a single equation in one unknown variable. To do this, note that the money market clearing condition implies that $M_t/(N_t p_t) = q^D(y, \Pi)$. Substituting this condition into (10.7) together with the fact that $\Pi^* = \mu/n$ yields:

$$g = \left[1 - \frac{1}{\mu^*}\right] q^D\left(y, \frac{\mu^*}{n}\right). \tag{10.10}$$

This equation implicitly defines the equilibrium growth rate of the money supply μ^* that is consistent with: individual optimization; rational expectations; and government budget balance.

The left-hand-side of equation (10.10) represents the value of government spending (per young person). The right-hand-side of this equation represents the value of the resources extracted by way of an inflation tax (per young person). The term in the square brackets represents the *inflation tax rate*. Notice that for a constant money supply ($\mu = 1$), the inflation tax rate is zero. On the other hand, a positive growth rate in the money supply implies a positive inflation tax. The term q^D in equation (10.10) represents the *inflation tax base*. The greater the willingness on the part of individuals to hold the government's money, the greater the ability of the government to tax them. Notice that if individuals do not value fiat money ($q^D = 0$), the government cannot raise any seigniorage revenue. As with any tax revenue, the total revenue collected is the product of the tax rate and the tax base.

Let $S(\mu) \equiv (1 - 1/\mu)q^D(y, \mu/n)$; i.e., the amount of seigniorage revenue collected when money grows at rate μ. We already know that no seigniorage revenue is collected when the money supply is held constant; i.e., $S(1) = 0$. But what happens when the government expands the money supply at a moderate rate? If the demand for money reacts negatively to the higher inflation rate (a reasonable assumption), then there are two offsetting effects on the amount of seigniorage revenue collected. On the one hand, the higher inflation tax implies more seigniorage revenue. On the other hand, the higher inflation rate reduces the tax base. If the first effect dominates the second, seigniorage revenue will rise. If the second effect dominates the first, then seigniorage revenue will fall. For very high inflation rates, it is likely that the second effect will dominate the first. Figure 10.4 plots the seigniorage revenue function $S(\mu)$ and the equilibria that are possible for a given fiscal policy parameter g.

FIGURE 10.4
Seigniorage and the Laffer Curve

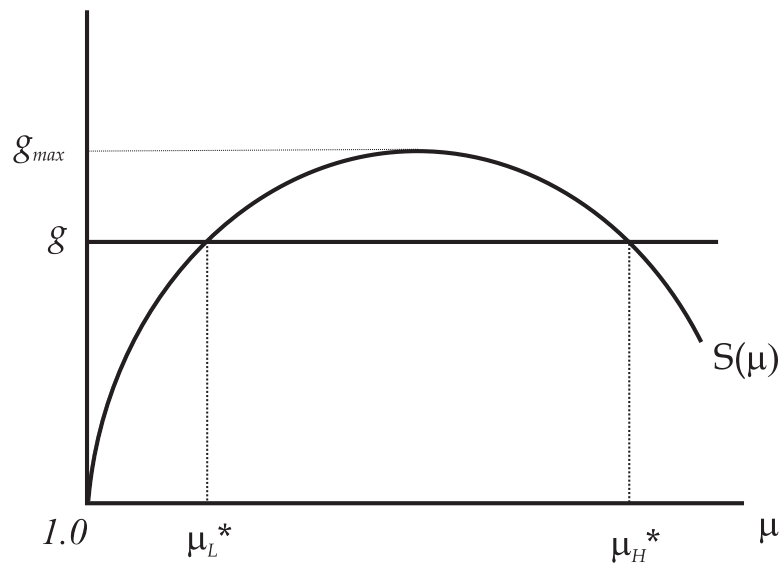


Figure 10.4 reveals that there is, in fact, a limit to a government's ability to raise revenue via an inflation tax. In particular, the amount of expenditure financed by printing money cannot exceed some finite number $g_{\max}$. The reason for this is because with ever higher rates of inflation, individuals begin to economize on their real money holdings, which reduces the tax base. The shape of the seigniorage tax function in Figure 10.4 resembles a 'Laffer Curve' (named after the economist, Art Laffer). The Laffer curve suggests that a government may actually collect more in the way of tax revenues by *decreasing* the tax rate (this type of argument is often heard among the so-called 'supply-side' economists). That is, while reducing the tax rate may reduce tax revenue, the resulting expansion in the tax base may more than make up for the decrease in the tax rate.

Figure 10.4 also reveals that for some given level of government spending $g < g_{\max}$, there are *two* equilibria that constitute possible outcomes. One of these equilibria is a 'low-inflation' regime (μ_L^*) and the other is a 'high-inflation' regime (μ_H^*). One can demonstrate that the utility of all individuals is higher under the low-inflation regime.

Finally, assuming that an economy is on the left-hand-side of the Laffer curve (the low-inflation regime), the model suggests that an expansion in government spending financed by printing fiat money is inflationary. This prediction appears to be broadly consistent with the historical evidence on fiscal crises.

10.3.2 The Inefficiency of Inflationary Finance

Like most taxes, the inflation tax distorts economic incentives and hence is a source of some economic inefficiency. What exactly is the nature of this inefficiency? Once again, we can appeal to our model to help identify the impact of inflation on economic activity and welfare.

Consider a government that requires g units of output (per young person). It is useful to consider first what would happen if the government chose to finance this expenditure with a lump-sum tax $\tau = g$. Lump-sum taxes are rare in reality, but it serves as a useful benchmark here since lump-sum taxes are non-distortionary.

With a lump-sum tax, the choice problem facing a young individual is to maximize $u(c_1, c_2)$ subject to $c_1 + \Pi c_2 = y - g$. Since the money supply is constant, we know that the equilibrium inflation rate is equal to $\Pi^* = 1/n$. Thus, the equilibrium budget constraint will be given by:

$$c_1 + \frac{c_2}{n} = y - g. \tag{10.11}$$

An increase in g serves to reduce the demand for both c_1 and c_2. This is just a pure wealth effect: the reduction in after-tax wealth reduces the demand for all normal goods; i.e., see Chapter 3. A lump-sum tax is non-distortionary because it does not alter the relative price of c_1 and c_2. Depict this scenario as point A in Figure 10.5.

Now imagine instead that the government chooses to finance its expenditures entirely by printing money. In this case, individuals face the following equilibrium budget constraint:

$$c_1 + \frac{\mu^*}{n} c_2 = y; \tag{10.12}$$

where μ^* is defined by equation (10.10). The equilibrium inflation rate here is given by $\Pi^* = \mu^*/n > 1/n$. The higher inflation rate distorts the relative price of c_1 and c_2 (making the 'cash' good more expensive relative to the 'non-cash' good).

How does the consumption allocation (c_1, c_2) differ across these two cases? Let (c_1^a, c_2^a) denote the allocation associated with the lump-sum tax. This allocation must satisfy:

$$MRS(c_1^a, c_2^a) = n;$$
$$c_1^a + \frac{c_2^a}{n} = y - g.$$

Let (c_1^b, c_2^b) denote the consumption allocation associated with the inflation tax.

This allocation must satisfy:

$$MRS(c_1^b, c_2^b) = \frac{n}{\mu^*};$$
$$c_1^b + \frac{\mu^*}{n} c_2^b = y.$$

The first thing we can establish is that (c_1^a, c_2^a) and (c_1^b, c_2^b) lie on a point at which the two budget lines cross. To see this, begin by rewriting the second budget constraint as follows:

$$\frac{\mu^*}{n} c_2^b = y - c_1^b = q^b.$$

Now, from equation (10.10), we see that $q^b = \left(\frac{\mu}{\mu - 1}\right) g$. Substituting this expression in the equation above then yields:

$$\frac{\mu^*}{n} c_2^b = \left(\frac{\mu^*}{\mu^* - 1}\right) g;$$

which implies that:

$$g = (\mu^* - 1) \frac{c_2^b}{n}. \tag{10.13}$$

Now, consider the second budget constraint again and manipulate it as follows:

$$c_1^b + \frac{\mu^*}{n} c_2^b + \frac{c_2^b}{n} = y + \frac{c_2^b}{n};$$
$$c_1^b + \frac{c_2^b}{n} = y - (\mu^* - 1) \frac{c_2^b}{n}.$$

Equation (10.13) shows that $(\mu^* - 1) c_2^b = g$; inserting this into the equation above yields:

$$c_1^b + \frac{c_2^b}{n} = y - g.$$

This establishes the fact that both (c_1^a, c_2^a) and (c_1^b, c_2^b) satisfy the condition $c_1 + c_2/n = y - g$.

So far, we have established that the two budget constraints intersect each other. But where do they intersect? Consider Figure 10.5. Do they intersect at a point like C (to the'left' of point A) or at a point like B (to the 'right' of point A)? To answer this question, we can appeal to the fact that:

$$MRS(c_1^a, c_2^a) = n > \frac{n}{\mu^*} = MRS(c_1^b, c_2^b).$$

That is, the slope of the indifference curve at (c_1^b, c_2^b) is 'flatter' than the slope at (c_1^a, c_2^a). This can only be true at a point like B.

FIGURE 10.5
Inflation Tax Distortion

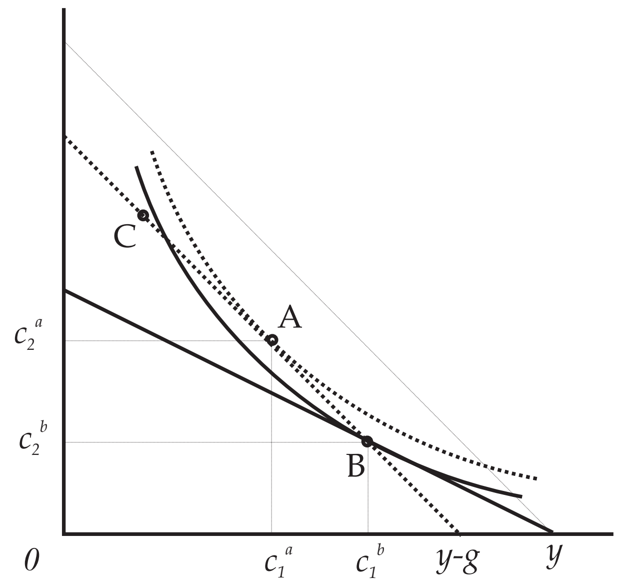


Figure 10.5 reveals the distortionary impact of financing government expenditures via an inflation tax. The higher inflation distorts the relative price of 'cash' goods (c_2) versus 'non-cash' goods (c_1), making the former more expensive relative to the latter. The resulting substitution away from cash goods lowers welfare (i.e., the indifference curve moves from point A to point B).

It is important to keep in mind here that an inflation tax is inefficient only relative to the largely hypothetical case in which a government has access to a lump-sum tax instrument. In reality, lump-sum taxes are often politically infeasible. Furthermore, in many developing countries, the bureaucratic infrastructure for collecting taxes may not be well-developed. In the developing world, much if not most economic activity still occurs in remote rural areas. Furthermore, even in urban areas, much economic activity may occur in 'informal' or underground markets, which are difficult to tax directly. In these circumstances then, an inflation tax may prove to be a relatively efficient way to collect tax revenue (even if it is inefficient relative to an hypothetical lump-sum tax).

- **Exercise 10.4**. In the model developed above, inflation does not affect the level of real GDP since the level of production is determined exogenously. However, imagine extending the model so that individuals can divide their time between work and leisure (as in Chapter 2). Leisure, or home-production more generally, is a good example of a 'non-cash' good. How would an inflation tax influence the level of real GDP in such an economy? Explain.

10.4 Summary

In most economies today, governments maintain a monopoly control over the supply of small denomination paper notes. Unlike in the past, these paper notes are fiat in nature; i.e., they are not backed by gold or any other real asset. Nevertheless, the availability of fiat money can improve economic efficiency by facilitating trades that might otherwise not occur (owing to a lack of double coincidence of wants and a lack of a public record-keeping technology). But since fiat money is intrinsically worthless, its value depends crucially on a self-fulfilling expectation. In particular, fiat money can only have value if people are confident that it will be valued.

The ability to print fiat money confers to the government an additional source of revenue called seigniorage. The ability to freely print fiat money does not, however, imply a limitless source of revenue for a government. Seigniorage is simply a tax, albeit an indirect tax, that reduces the purchasing power (transferring it to the government) of all individuals who hold it. An important limit to collecting an inflation tax is given by the willingness and ability of individuals to substitute out of activities that require the use of fiat money. In an open economy, a further limit to seigniorage may be imposed by the willingness and ability of individuals to substitute out of the domestic currency into other currencies.

While an inflation tax may generate some economic inefficiency, it may nevertheless constitute a relatively efficient way to collect at least some taxes. This is particularly true of lesser-developed economies that lack an efficient method for collecting taxes directly.

10.5 References

1. Kocherlakota, Narayana (1998). "The Technological Role of Fiat Money," *Federal Reserve Bank of Minneapolis Quarterly Review*, Summer, pp. 2–10.

2. Samuelson, Paul A. (1958). "An Exact Consumption-Loan Model of Interest with or without the Social Contrivance of Money," *Journal of Political Economy*, 66: 467–82.

3. Smith, Bruce D. (1985). "American Colonial Regimes: The Failure of the Quantity Theory and Some Evidence in Favour of an Alternative View," *Canadian Journal of Economics*, 18: 531–65.

Chapter 11

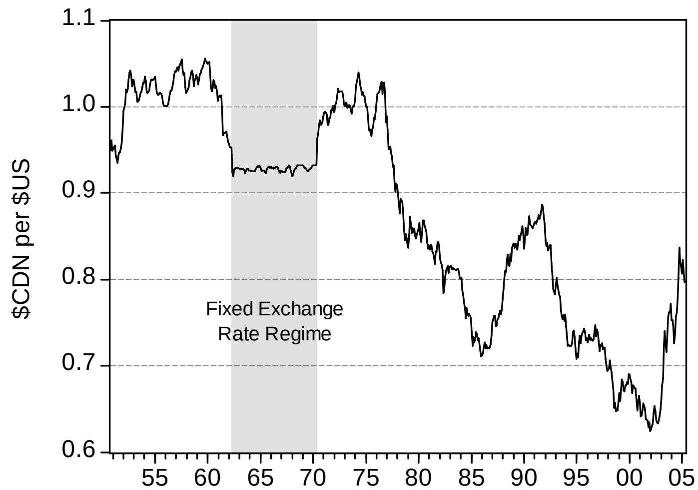
International Monetary Systems

11.1 Introduction

Almost every day we are presented with news concerning the behavior of the nominal exchange rate. What is a nominal exchange rate? A nominal exchange rate is simply the relative price of two currencies at a given point in time. For example, the nominal exchange rate between the Canadian dollar (CDN) and the U.S. dollar (USD) currently stands around 0.80. What this means is that one can currently purchase $1 CDN for $0.80 USD on the foreign exchange market. Alternatively, one can purchase $1/0.8 = 1.20$ $CDN for one $USD. Figure 11.1 plots the Canada-U.S. exchange rate since 1950 (USD per CDN).

Following the end of the second world war, many countries agreed to fix their exchange rates in terms of USD, with the USD itself fixed to the price of gold (one ounce of gold worth $35 US). This system of fixed exchange rates was known as the Bretton-Woods agreement. (Canada joined the Bretton-Woods agreement in the early 1960s). The Bretton-Woods agreement was abandoned in the early 1970s when the United States abandoned its policy of pegging the USD to gold. Since the abandonment of the Bretton-Woods agreement, exchange rates have largely been left to 'float' (i.e., determined primarily by market forces). Many economists who advocated the benefits of a floating exchange rate regime have been surprised by what many regard to be an 'excessive' volatility in exchange rate behavior.

FIGURE 11.1
Canada - U.S. Exchange Rate



The perception of excessive exchange rate volatility has led many people to question the wisdom of a floating exchange rate system. In fact, some countries have gone so far as to abandon their national currencies in favor of a common regional currency (e.g., the European Currency Union). Other countries, like Argentina during the 1990s, attempted to unilaterally fix their exchange rate relative to the USD (this experiment was abandoned in 2002). Yet other countries, such as Panama, have essentially abandoned their local currency and have 'dollarized' (i.e., adopted the USD).[1]

What are the market forces that determine the nominal exchange rate between two currencies? Is there any reason to believe that foreign exchange rate markets are 'different' from other markets? Is there any reason to believe that market-determined exchange rates are likely to display 'excessive' volatility driven largely by 'speculative' forces? Is there any rationale in adopting a multilateral fixed exchange rate regime, like the Bretton-Woods arrangement? Is there any rationale for a country to embark on a policy of unilaterally fixing its exchange rate relative to some other currency? Is there any rationale for a country to 'dollarize' or to arrange for a common currency among a group of major trading partners (such as the Euro)? These are all interesting questions. In what follows, the theory developed in the previous chapter will be brought to bear on these questions.

[1] Panamanian currency (called the Balboa) exists only in the form of coins. These coins trade at a fixed exchange rate to the USD. All paper money in Panama consist of USD.

11.2 Nominal Exchange Rate Determination: Free Markets

If all government monies were fully backed by real assets (like gold), then there would be no reason to think of foreign exchange markets as being 'different' from the market for any other commodity. For example, suppose that the Canadian dollar was made redeemable for one ounce of gold and suppose that the U.S. dollar was made redeemable for two ounces of gold. Then logic (i.e., a no-arbitrage condition) suggests that one Canadian dollar should be able to purchase 0.5 U.S. dollars.

However, government monies today are not backed in any way; i.e., they are fiat in nature. Fiat monies are, by definition, intrinsically useless objects created by governments. This fact alone suggests that foreign exchange markets may indeed be 'different' from other types of markets (which involve either an exchange of goods, or claims to goods). What are the market forces that determine the exchange rate between two intrinsically useless objects? To answer this question, consider the following model.

Imagine a world consisting of two countries a and b. Each country is populated by overlapping generations of individuals that live for two periods. Young individuals have a nonstorable endowment $y > 0$. This endowment is the same for young individuals across countries and consists of the same output (i.e., the output of country a is indistinguishable from the output of country b). This latter assumption effectively fixes the *real* exchange rate between country a output and country b output at unity. Such a simplification allows us to focus on the forces that determine the nominal exchange rate.

Assume that individuals may freely trade goods internationally. In particular, what this means is that a young agent in country a may freely sell his output to an old agent in country b; and vice-versa.

It will be useful to simplify preferences such that people care only for consumption when old; i.e., $U(c_1, c_2) = u(c_2)$. None of our conclusions will depend on this assumption. The assumption simplifies matters because it implies that young individuals will find it optimal to save their entire endowment. Since saving in this model can only occur in the form of fiat money holdings, this assumption essentially fixes the demand for real money balances; i.e., $q^D = y$. Assume that individuals in both countries have the same preferences. As well, and again for simplicity only, assume that the population in each country is fixed at N.

Each country has its own national money M^a and M^b. Assume that the money supply in each country remains fixed over time. Let p_t^a and p_t^b denote the price-level in country a and b, respectively. Since there is free-trade in output across countries, a simple no-arbitrage condition implies that the following condition must hold:

$$p_t^a = e_t p_t^b; \tag{11.1}$$

where e_t denotes the price of country b money measured in units of country a money (i.e., the nominal exchange rate). Equation (11.1) is sometimes called the *Law of One Price* (LOP). The LOP states that any good or asset that is freely traded across two countries should sell at the same price (abstracting from transportation costs, taxes, etc.), once the proper exchange rate conversion is made. That is, suppose that you can purchase one share of General Motors on the New York Stock Exchange for \$34.45 USD and the same share for \$43.34 CDN on the Toronto Stock Exchange (these are the respective closing prices as of June 13, 2005). Then the LOP suggests that the Canada-U.S. nominal exchange rate should be equal to $e_t = 34.45/43.34 \approx 0.795$. (In fact, this turned out to be the closing rate on June 13, 2005). More generally then, the LOP suggests that:

$$e_t = \frac{p_t^a}{p_t^b}.$$

Now, since individuals in this model have a fixed demand for real money balances $q^D = y$, their desired *total* money holdings are independent of the rate of return on money. However, if individuals are free to hold and transact in either money, their *relative* rates of return will be important for determining the *composition* of an individual's portfolio of money holdings; i.e., $q_a^D + q_b^D = y$. The rate of return on each money is inversely related to the inflation rate in each country; i.e.,

$$\Pi^a \equiv \frac{p_{t+1}^a}{p_t^a} \text{ and } \Pi^b \equiv \frac{p_{t+1}^b}{p_t^b}.$$

Let us abstract from any uncertainty, so that inflation rates are deterministic. In this case, we can appeal to a simple no-arbitrage condition that states if individuals are to willingly hold both monies in their wealth portfolios, each money must earn an identical rate of return; i.e.,[2]

$$\Pi^a = \Pi^b = \Pi.$$

Given that both monies earn the same rate of return, and given that individuals are free to hold and transact in either money, how much of each money should individuals hold? The answer is that individuals should not care about the composition of their money holdings; i.e., the individual demands for each money are *indeterminate* under these circumstances.

In the world described above, the monies of country a and country b are viewed as *perfect substitutes*. In other words, there is no independent money market for each money. There is only a single world supply and world demand for money; i.e., the relevant money-market clearing condition is given by:

$$\frac{M^a}{p_t^a} + \frac{M^b}{p_t^b} = 2N\left(q_a^D + q_b^D\right).$$

[2] In the case of uncertainty, one can show that the *expected* rates of return on each money must be equated with the inflation rate in each country restricted only to follow a martingale; i.e., see Appendix 11.A.

The left-hand-side of this equation represents the total world supply of real money balances, while the right-hand-side describes the total world demand for real money balances. Using the fact that $e = p_t^a/p_t^b$ and $q_a^D + q_b^D = y$, this equation may be rewritten as:

$$M^a + e_t M^b = p_t^a 2Ny. \tag{11.2}$$

Equation (11.2) constitutes one equation in *two* unknowns: e_t and p_t^a. Obviously, there are infinite combinations of e_t and p_t^a that satisfy this restriction. To determine the exchange rate (and price-level), we need another equation. However, economic theory does not deliver any further restrictions in this environment. In other words, there are no economic fundamentals that determine the nominal exchange rate. *Any* nominal exchange rate e_t (together with a corresponding price-level that satisfies 11.2) is consistent with an equilibrium. *Which* exchange rate actually prevails can depend entirely on 'non-fundamental' forces, like self-fulfilling market expectations. To put things another way, the equilibrium nominal exchange rate between two fiat currencies is driven *entirely by market speculation.*

11.2.1 Understanding Nominal Exchange Rate Indeterminacy

To many people, especially those with great faith in the institution of free markets, the idea that free markets are incapable of determining the relative price of two objects is somewhat of a puzzle. But the key to understanding this puzzle is that fiat monies are not like any other goods or assets. Fiat currencies are intrinsically worthless objects; it should come as no surprise that a free market is incapable of determining a 'fundamental' relative price of two intrinsically worthless objects.

Let me try to provide some intuition for this indeterminacy result by developing a series of examples. Imagine that you are sitting down with your friends to play poker. Your host brings out the poker chips. Poker chips are intrinsically useless objects; they are distinguished from each other only by color (red, white and blue). The group must decide on an exchange rate system for poker chips. Are there any *fundamental* market forces that determine what these exchange rates should be?

Alright, let's consider another example. Imagine that the Bank of Canada prints up two types of paper notes: One note is blue (and is called a *Laurier*), while the other note is green (and is called a *Queen*). Imagine that the Bank of Canada neglects to place any numbers on the notes, choosing instead to let a free market in these notes determine the rate at which they exchange. Can you identify any fundamental market forces that would pin down a precise exchange rate between Lauriers and Queens?

Imagine now that there are two fiat currencies called Lincolns and Lauriers.

Imagine that Lincolns are printed by the Federal Reserve Bank of the United States. Imagine further that both Lincolns and Lauriers are associated with the number five (as they indeed are in reality). In many towns along the Canada-U.S. border, merchants view Lincolns and Lauriers as perfect substitutes (i.e., they are willing to accept either USD or CDN as payment). If these two fiat monies are viewed as perfect substitutes, then what, if anything, determines the rate at which they exchange for each other?

In each of these examples, it should be clear that if different fiat monies are viewed as perfect substitutes for each other, then it is difficult to identify any market forces that would pin down their exchange rates.

In Canada, the nominal exchange rate between Lauriers and Queens is four-to-one. Likewise, the nickels and dimes can be exchanged for two-to-one. How were these nominal exchange rates determined? Is a Queen worth four times more than a Laurier because she is prettier, or is from a royal family? And why does the nominal exchange rate between Queens and Lauriers remain constant over time? Are there separate and stable supplies and demands for Queens and Lauriers?

No, of course not. The system of nominal exchange rates between different types of Canadian money is determined by the Bank of Canada. Lauriers and Queens exchange at four-to-one because the Bank of Canada (which prints these notes) stands ready to exchange these two notes at the stated rate (as indicated by the number 5 on the Laurier and the number 20 on the Queen).

Let us now reconsider equation (11.2), but in the context of a Canadian economy that is closed to international trade. Let M^a denote the supply of Lauriers and let M^b denote the supply of Queens. Thus, p_t^a now represents the price of output measured in Lauriers and p_t^b measures the price of output measured in Queens. Suppose that the Bank of Canada sets the nominal exchange rate to $e = 4$, so that one Queen is worth four Lauriers. Then clearly, $p_t^a = 4p_t^b$. In other words, you have to sacrifice four times as many Lauriers (relative to Queens) to purchase the same good. Now, with the nominal exchange rate fixed in this manner, note that condition (11.2) constitutes one equation in the one unknown; i.e., p_t^a. Thus, a government policy that fixes the nominal exchange rate implies that market forces can determine a unique price-level p_t^a (with $p_t^b = 0.25p_t^a$).

Perhaps you are surprised to learn that most countries maintain a system of *fixed exchange rates* for their own national monies (i.e., perhaps you've never thought of nominal exchange rates in quite this way before). But the fundamental difference between Lauriers and Queens is as absent as it is between Lauriers and Lincolns; i.e., they are just intrinsically useless bits of paper. Understanding this leads to a natural question: If fixing the nominal exchange rate between monies within a nation makes sense, why does it also not make sense to fix the nominal exchange rate between monies across nations? In fact, why even bother with different national currencies? Why do nations simply not agree to adopt a single common currency?

11.2.2 A Multilateral Fixed Exchange Rate Regime

If nominal exchange rate fluctuations are driven primarily by speculation (exogenous shifts in expectations that are not based on fundamentals), then the resulting uncertainty is likely welfare-reducing.[3] In principle, nominal exchange rate uncertainty would pose little concern if individuals held their wealth in a diversified portfolio consisting of assets denominated in all the world's currencies. But for various reasons, individuals do not appear to behave in this manner. Likewise, individuals could, in principle, try to hedge foreign exchange risk by purchasing insurance. In fact, many companies do behave in this manner. But many do not and, in any case, hedging is costly. Many economists believe that international trade would be facilitated (and welfare improved) in the absence of nominal exchange rate risk. One way to eliminate such risk is to enter into a multilateral agreement with other countries to fix the exchange rate. The Bretton-Woods arrangement (1946–1971) is a classic example of such an agreement.[4]

A multilateral fixed exchange rate regime sounds like a great idea–in principle. However, as you may have guessed from the collapse of the Bretton-Woods arrangement, fixed exchange rate regimes are not without their problems. One of the fundamental problems with maintaining a fixed exchange system is that it requires a high degree of coordination between countries in how fiscal and monetary policies are to be conducted. To see why this is so, let us combine elements of the model developed above with what we have learned earlier about the use of money as a revenue device.

Let us consider a bilateral fixed exchange rate agreement between two countries a and b. The two countries differ in their fiscal policies. In particular, the government in country a has a 'large' fiscal spending program $g > 0$, whereas country b has a 'small' fiscal spending program (that we can normalize to zero). Imagine that both countries agree to fix the exchange rate between their two currencies at par, so that $e = 1$. In this case, the LOP implies that both countries should share the same price-level p_t.

Assume that the government in country b maintains the supply of its currency at some constant level M^b. Presumably, country b expects country a to maintain its money supply at some constant level as well. Note that if country a keeps its money supply constant, then it must finance its expenditures entirely by taxing its own citizens. If each government behaved in this manner, then a

[3] Some economists blame the speculative nature of free markets in creating excessive exchange rate volatility. In fact, the blame lies more with governments that insist on printing fiat currency, which has no intrinsic value and is therefore impossible to price. If governments were always well-behaved and if they backed their money with real assets, market forces would have no problem in determining exchange rates on the basis of fundamentals.

[4] See: http://en.wikipedia.org/ wiki/ Bretton_Woods_Conference

fixed exchange rate system should work well. But governments are not always so well-behaved.

In particular, imagine that the government in country a is under political pressure to keep spending high (e.g., perhaps it has a war to fight in Vietnam) and is under political pressure to keep taxes low. Then the fiscal authority in country a may find it irresistible to resort to the printing press to alleviate some of these fiscal pressures. If country a chooses to finance all government expenditures with money creation, then the government budget constraint implies:

$$\begin{aligned} p_t N g &= M_t^a - M_{t-1}^a; \\ &= \left[1 - \frac{1}{\mu_t^a}\right] M_t^a. \end{aligned}$$

If country a was closed to international trade (or if country b citizens were legally prohibited from holding country a money), then the domestic money-market clearing condition is given by:

$$M_t^a = p_t^a N y,$$

(since $q^D = y$ here). This condition implies that the equilibrium domestic inflation rate is given by $\mu_t^a = \mu^a$. Now, substitute the money market clearing condition into the government budget constraint:

$$\begin{aligned} g &= \left[1 - \frac{1}{\mu^a}\right] \frac{M_t^a}{N p_t^a}; \\ &= \left[1 - \frac{1}{\mu^a}\right] y. \end{aligned}$$

Thus, the domestic inflation rate that would be required to finance the expenditure level g is given by:

$$\mu^a = \frac{1}{1 - \frac{g}{y}} > 1. \qquad (11.3)$$

- **Exercise 11.1**. The ratio (g/y) represents the ratio of government purchases to GDP. Compute the domestic inflation rate that would be necessary to finance all of government spending if $(g/y) = 0.20$.

Let us now reconsider the calculations above under the assumption of free trade in goods and money across the two countries (with a fixed exchange rate $e = 1$). In this case, the government budget constraint in country a is given by:

$$\begin{aligned} p_t N g &= M_t - M_{t-1}; \\ &= \left[1 - \frac{1}{\mu}\right] M_t; \end{aligned}$$

where $M_t = M_t^a + eM^b$ denotes the *total* money supply across the two countries. Under free trade in goods and money, the relevant money-market clearing condition is given by:

$$\begin{aligned} M_t &= p_t N(q^a + q^b); \\ &= p_t N 2y; \end{aligned}$$

(since $q^a = q^b = y$ here). Now, substitute this money-market clearing condition into the government budget constraint:

$$\begin{aligned} g &= \left[1 - \frac{1}{\mu}\right] \frac{M_t}{Np_t}; \\ &= \left[1 - \frac{1}{\mu}\right] 2y. \end{aligned}$$

In this case, the inflation rate that would be required to finance the expenditure level g is given by:

$$\mu = \frac{1}{1 - \frac{1}{2}\frac{g}{y}}. \tag{11.4}$$

Now, compare equations (11.3) and (11.4). What this comparison tells us is that country a is able to finance the same level of government spending with a *lower* inflation rate under a fixed exchange rate system.

- **Exercise 11.2**. Redo Exercise 11.1 assuming a fixed exchange rate regime and free trade in goods and money (i.e., using equation (11.4) instead of equation (11.3).

What is going on here? Note that since the exchange rate is fixed at $e = 1$, it follows that the price level p_t is common across both countries. What this means is that the inflation generated by country a is 'exported' to country b. That is, the inflation rate calculated in equation (11.4) now applies to both countries. Thus, under a bilateral fixed exchange rate agreement, the tax base is effectively doubled for country a (since both countries are the same size). The government in country a can now finance the same level of government spending with a lower inflation rate because the tax base is so much larger. In other words, the residents of country b end up incurring half the bill for the government expenditures in country a.

To better understand this phenomenon, imagine that we have two monies called Queens and Lauriers. Queen's are worth \$20 and Lauriers are worth \$5. The exchange rate is fixed by government policy at four to one. Now imagine that the Bank of Canada starts printing large numbers of Queens, while holding the supply of Lauriers fixed. What do you think would happen to the exchange rate between Queens and Lauriers? The answer, of course, is nothing. What do you think would happen to the value of money (i.e., the price-level)? One would imagine that the price-level would rise (i.e., the value of money would fall). However, note that because the exchange rate is fixed between Queens and Lauriers, the expansion in the supply of Queens will reduce the value of *all* money in equal proportion.

This example illustrates one of the fundamental problems associated with maintaining a multilateral fixed exchange rate agreement. In order for such an agreement to work well, countries must agree to restrain their monetary and fiscal policies in an appropriate manner. Individual countries may try to 'cheat' a little bit by expanding the supply of their domestic currency and thereby export some of the inflation tax burden to other countries. If all countries behaved in this way, the result would be a high rate of world inflation. Or an individual country may find it difficult to restrain its monetary policy in times of fiscal crisis. This is arguably what happened in the United States in the late 1960s and early 1970s, as the fiscal authority struggled to meet the fiscal pressures building from the escalation of its war in Vietnam. Under the Bretton-Woods agreement, the United States was financing a part of its (widely unpopular) war effort through an inflation tax that was paid in part by all members of the Bretton-Woods agreement. Ultimately, the Bretton-Woods agreement collapsed because of these pressures.

- **Exercise 11.3**. The supply of fiat money in Canada is controlled by a single central agency (the Bank of Canada, located in Ottawa). Imagine instead that each province of Canada was given the authority to print Canadian dollars. What do you think might happen to the inflation rate? Explain.

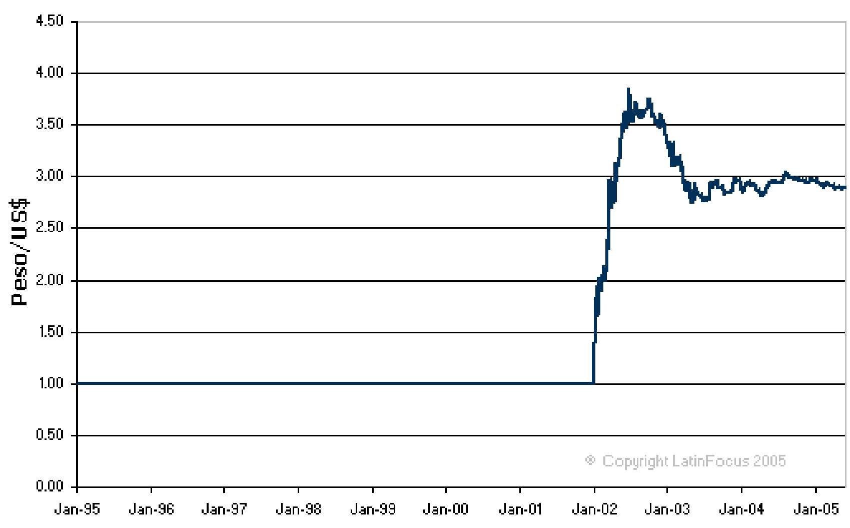
11.2.3 Speculative Attacks

The key to maintaining a fixed exchange rate system is obtain a *credible commitment* on the part of all member governments to exchange different monies at the stated rates. Such a commitment is necessary to defend the exchange rate system against speculative attacks. To see why, consider the following scenario. Imagine that Canada and the U.S. have entered into a bilateral fixed exchange rate agreement. Now, imagine that participants in the foreign exchange (FX) market 'speculate' that the Canadian dollar may depreciate. Such a speculation may lead market participants to 'dump' their Canadian dollars on the FX market (in exchange for U.S. dollars). But *if* the U.S. government stands ready to print all the U.S. dollars demanded by speculators (in exchange for Canadian dollars) at the stated rate, then the speculative attack *must* fail. Understanding that this must be the case, there is no reason to engage in such speculative activity.

However, many countries attempt to fix their exchange rate *unilaterally*, often by way of a currency board. For example, Argentina adopted a currency board from April 1, 1991 through January 6, 2002 with the stated intent of fixing the value of its Peso to the U.S. dollar at par. Defending one's exchange rate against speculative attacks is more difficult to do unilaterally than it is via a bilateral agreement. This is because the commitment to defend the exchange rate must rest solely on the country imposing the peg. In particular, the United States did not promise to help Argentina defend the Peso in the event of a speculative attack. To defend its currency unilaterally, Argentina had to convince FX participants that it stood ready to do whatever it took to maintain the exchange rate. One way to do this is for the currency board to hold one U.S. dollar in reserve for every Peso it prints (this reserve currency must ultimately be acquired via taxation, if the Fed has no desire to hold Pesos). Alternatively, the Argentine government must stand willing to tax its citizens to acquire the U.S. dollars it needs to meet the demands of any speculators. More importantly, FX participants must *believe* that the Argentine government would be willing to take such an action; i.e., the stated policy must be perceived to be *credible*.

Figure 11.2 plots the exchange rate between the Argentine Peso and the USD for 1995-2005. For a period of time, the Argentine currency board appeared to work well, at least, in terms of maintaining a fixed rate of exchange (par) with the USD. However, for a variety of reasons, the currency board was compelled to abandon its peg against the USD in January 2002. Figure 11.2 shows that following this abandonment, the Argentine Peso devalued sharply and is presently worth around $0.34 USD (about one-third of its former value).

FIGURE 11.2
Argentine Peso per USD (1995–2005)



What went wrong in Argentina? According to my Argentine friends, nothing went wrong–what happened was perfectly normal (which is to say that everything is always going wrong in Argentina). Some people place the 'blame' on the U.S. dollar, which strengthened relative to most currencies over the 1990s. Since the Peso was linked to the U.S. dollar, this had the effect of strengthening the Peso as well, which evidently had the effect of making Argentina's exports uncompetitive on world markets. While there may be an element of truth to this argument, one wonders how the U.S. economy managed to cope with the rising value of its currency over the same period (in which the U.S. economy boomed). Likewise, if the rising U.S. dollar made Argentine exports less competitive, what prevented Argentine exporters from cutting their prices?

A more plausible explanation may be the following. First, the charter governing Argentina's currency board did not require that Pesos be fully backed by USD. Initially, as much as one-third of Pesos issued could be backed by Argentine government bonds (which are simply claims to future Pesos). In the event of a major speculative attack, the currency board would not have enough USD reserves to defend the exchange rate. Furthermore, it would likely have been viewed as implausible to expect the Argentine government to tax its citizens to make up for any shortfall in reserves. Second, a combination of a weak economy and liberal government spending led to massive budget deficits in the late 1990s. The climbing deficit led to an increase in devaluation concerns. According to Spiegel (2002), roughly $20 billion in capital 'fled' the country in 2001.[5] Market participants were clearly worried about the government's ability to finance its growing debt position without resorting to an inflation tax (Peso interest rates climbed to between 40-60% at this time). In an attempt to stem the outflow of capital, the government froze bank deposits, which precipitated a financial crisis. Finally, the government simply gave up any pretense concerning its willingness and/or ability to defend the exchange rate. Of course, this simply served to confirm market speculation.

At the end of the day, the currency board was simply not structured in a way that would allow it to make good on its promise to redeem Pesos for USD at par. In the absence of full credibility, a unilateral exchange rate peg is an inviting target for currency speculators.

- **Exercise 11.4**. Explain why speculating against a currency that is pegged unilaterally to a major currency like the USD is close to a 'no-lose' betting situation. Hint: explain what a speculator is likely to lose/gain in either scenario: (a) a speculative attack fails to materialize; and (b) a speculative attack that succeeds in devaluing the currency.

[5] I presume what this means is that Argentines flocked to dispose of $20 billion in Peso-denominated assets, using the proceeds to purchase foreign (primarily U.S.) assets.

11.2.4 Currency Union

A currency union is very much like a multilateral fixed exchange rate regime. That is, different monies with fixed nominal exchange rates essentially constitute a single money. The only substantive difference is that in a currency union, the control of the money supply is taken out of the hands of individual member countries and relegated to a central authority. The central bank of the European Currency Union (ECU), for example, is located in Frankfurt, Germany, and is called the European Central Bank (ECB). The ECB is governed by a board of directors, headed by a president and consisting of the board of directors and representatives of other central banks in the ECU. These other central banks now behave more like the regional offices of the Federal Reserve system in the United States (i.e., they no longer exert independent influence on domestic monetary policy).

Having a centralized monetary authority is a good way to mitigate the lack of coordination in domestic monetary policies that may potentially afflict a multilateral fixed exchange rate system. However, as the recent European experience reveals, such a system is not free of political pressure. In particular, ECB members often feel that the central authority neglects the 'special' concerns of their respective countries. There is also the issue of how much seigniorage revenue to collect and distribute among member states. The governments of member countries may have an incentive to issue large amounts of nominal government debt and then lobby the ECB for high inflation to reduce the domestic tax burden (spreading the tax burden across member countries). The success of a currency union depends largely on the ability of the central authority to deal with a variety of competing political interests. This is why a currency union within a country is likely to be more successful than a currency union consisting of different nations (the difference, however, is only a matter of degree).

11.2.5 Dollarization

One way to eliminate nominal exchange rate risk that may exist with a major trading partner is to simply adopt the currency of your partner. As mentioned earlier, this is a policy that has been adopted by Panama, which has adopted the U.S. dollar as its primary medium of exchange. Following the long slide in the value of the Canadian dollar since the mid 1970s (see Figure 11.1), many economists were advocating that Canada should adopt a similar policy.

One of the obvious implications of adopting the currency of foreign country is that the domestic country loses all control of its monetary policy. Depending on circumstances, this may be viewed as either a good or bad thing. It is likely a good thing if the government of the domestic country cannot be trusted to maintain a 'sound' monetary policy. Any loss in seigniorage revenue may be more than offset by the gains associated with a stable currency and no exchange rate risk. On the other hand, should the foreign government find itself in a fiscal crisis, the value of the foreign currency may fall precipitously through an unexpected inflation. In such an event, the domestic country would in effect be helping the foreign government resolve its fiscal crisis (through an inflation tax).

- **Exercise 11.5**. If the Argentine government had simply dollarized instead of erecting a currency board, would a financial crisis have been averted? Discuss.

11.3 Nominal Exchange Rate Determination: Legal Restrictions

The previous section described a world in which individuals are free to trade internationally and free to hold different types of fiat monies. Since fiat money is an intrinsically useless object, one fiat money is as good as any other fiat money; i.e., in such world, different fiat monies are likely to be viewed as perfect substitutes for each other. But if this is true, then there are no market forces that pin down a unique exchange rate system between different fiat monies: the nominal exchange rate is indeterminate. This indeterminacy problem can be resolved only by government policy; i.e., via membership in a multilateral exchange arrangement, or via the adoption of a common currency.

The world so described appears to ring true along many dimensions. In particular, it seems capable of explaining why market-determined exchange rates appear to display 'excessive' volatility. And it also explains why governments are often eager to enter into multilateral fixed exchange rate arrangements. But this view of the world is perhaps too extreme. In particular, if world currencies are indeed perfect substitutes, then one would expect the currencies of different countries to circulate widely within national borders. Casual observations suggests, however, that national borders do, in large measure, determine currency usage. Furthermore, it is difficult (although, not impossible) to reconcile the indeterminacy proposition with many historical episodes in which exchange rates have floated with relative stability (see, for example, the behavior of the Canada-U.S. exchange rate in the 1950s in Figure 11.1).

One element of reality that is missing from the model developed above is the absence of legal restrictions on money holdings. These types of legal restrictions are called *foreign currency controls* (FCCs). Foreign currency controls come in a variety of guises. For example, chartered banks are usually required to hold reserves of currency consisting primarily of domestic money or a restricted from offering deposits denominated in foreign currencies.[6] Many countries have 'capital controls' in place that restrict domestic agents from undertaking capital account transactions with foreign agents in an attempt to keep trade 'balanced'

[6] In the late 1970s, the Bank of America wanted to offer deposits denominated in Japanese yen, but was officially discouraged from doing so.

(i.e., to reduce a growing current account deficit). An example of such a capital control is a restriction on the ownership of assets not located in the country of residence. In some countries, more Draconian measures are imposed; e.g., legal restrictions are imposed that prohibit domestic residents from holding any foreign money whatsoever. Such legal restrictions, whether current or anticipated, have the effect of generating a well-defined demand for individual currencies. If the demands for individual currencies become well-defined in this manner, then nominal exchange rate indeterminacy may disappear.

To see how this might work, let us consider an example that constitutes the opposite extreme of the model studied above. Let us again consider two countries, labelled a and b. It will be helpful to generalize the analysis here to consider different population growth rates n^i and different money supply growth rates μ^i for $i = a, b$.

Now, imagine that the governments in each country impose foreign currency controls. Assume that this legal restriction does not prohibit international trade (so that the young in one country may still sell output to the old of the other country). But the legal restriction prohibits young individuals from carrying foreign currency from one period to the next (i.e., domestic agents can only save by accumulating domestic currency). In this case, if a young agent from country a meets an old agent from country b, the young agent may 'export' output to country b in exchange for foreign currency. But the FCC restriction requires that the young agent in possession of the foreign currency dispose of it within the period on the foreign exchange market (in exchange for domestic currency).

The effect of these FCCs is to create two separate money markets: one for currency a and one for currency b. In other words, each country now has its own money supply and demand that independently determine the value of its fiat money; i.e.,

$$M_t^a = p_t^a N_t^a y;$$
$$M_t^b = p_t^b N_t^b y.$$

With domestic price-levels determined in this way, the equilibrium exchange is determined (by the LOP) as:

$$e_t = \frac{p_t^a}{p_t^b} = \frac{M_t^a}{M_t^b}\frac{N_t^b}{N_t^b}. \tag{11.5}$$

The equilibrium inflation rate in each country (the inverse of the rate of return on fiat money) is now determined entirely by domestic considerations; i.e.,

$$\Pi^a = \frac{\mu^a}{n^a} \text{ and } \Pi^b = \frac{\mu^b}{n^b} \ .$$

From this, it follows that the time-path of the equilibrium exchange rate must follow:

$$\frac{e_{t+1}}{e_t} = \frac{\Pi^a}{\Pi^b} = \frac{\mu^a}{\mu^b}\frac{n^b}{n^a}. \tag{11.6}$$

Thus, in the presence of such legal restrictions, the theory predicts that if exchange rates are allowed to float, they will be determined by the relative supplies and demands for each currency. Equation (11.5) tells us that, holding all else constant, an increase in the supply of country a money will lead to a depreciation in the exchange rate (i.e., e_t, which measures the value of country b money in units of country a money, rises). Equation (11.6) tells us that, holding all else constant, an increase in the growth rate of country a money will cause it to appreciate in value at a slower rate (and possibly depreciate, if $e_{t+1}/e_t < 1$).

11.3.1 Fixing the Exchange Rate Unilaterally

Under a system of foreign currency controls, a country can in principle fix the exchange rate by adopting a simple monetary policy. Equation (11.6) suggests how this may be done. A fixed exchange rate implies that $e_{t+1} = e_t$. This implies that country a could fix the exchange rate simply by setting its monetary policy to satisfy:

$$\mu^a = \frac{n^b}{n^a}\mu^b.$$

Essentially, this policy suggests that country a monetary policy should follow country b monetary policy. In other words, this model suggests that a country can choose either to fix the exchange rate or to pursue an independent monetary, *but not both simultaneously.*

11.4 Summary

Because foreign exchange markets deal with the exchange of intrinsically useless objects (fiat currencies), there is little reason to expect a free market in international monies to function in any well-behaved manner. In the absence of legal restrictions, or other frictions, one fiat object has the same intrinsic value as any other fiat object (zero). Free markets are good at pricing objects with intrinsic value; there is no obvious way to price one fiat object relative to another. It is too much to ask markets to do the impossible.

If governments insist on monopolizing small paper note issue with fiat money, how should international exchange markets be organized? One possible answer to this question is to be found in the way nations organize their internal money markets. Most nations delegate the creation of fiat money to a centralized institution (like a central bank). In particular, cities, provinces and states within a nation are not free to pursue distinct seigniorage policies. The different monies that circulate within a nation trade at fixed exchange rates (creating different denominations) that are determined by the monetary authority. By and large, this type of system appears to work tolerably well (most of the time) in a relatively politically integrated structure like a nation.

Does the same logic extend to the case of a world economy? Imagine a world with a single currency. People travelling to foreign countries would never have to first visit the foreign exchange booth at the airport. Firms engaged in international trade could quote their prices (and accept payment) in terms of a single currency. No one would ever have to worry about foreign exchange risk. Such a world is *theoretically* possible. But such an arrangement would have to overcome several severe political obstacles. First, a single world currency would require that nations surrender their sovereignty over monetary policy to some trusted international institution. (Given the dysfunctional nature of the U.N. and the IMF, one may legitimately question the feasibility of this requirement alone). This centralized authority would have to settle on a 'one-size-fits-all' monetary policy and deal with the politically delicate question of how to distribute seigniorage revenue 'fairly.' Given that there are significant differences in the extent to which international governments rely on seigniorage revenue, reaching a consensus on this matter seems highly unlikely.

If a single world currency is politically infeasible, a close 'next-best' alternative would be a multilateral fixed exchange rate arrangement, like Bretton-Woods (*sans* foreign currency controls). Under this scenario, different currencies function as different denominations of the world money supply, freely traded everywhere. Such a regime requires a high degree of coordination among national monetary policies in order to prevent speculative attacks. More importantly, it requires significant restraint on the part of national treasuries from pressuring the local monetary authorities into 'monetizing' local government debt. Under such a system, the temptation to export inflation to other countries may prove to be politically irresistible. This type of political pressure is likely behind the collapse of every international fixed exchange rate system ever devised (including Bretton-Woods).

If common currency and multilateral exchange rate arrangements are both ruled out, then another alternative would be to impose foreign currency controls and allow the market to determine the exchange rate. But while foreign currency controls eliminate the speculative dimension of exchange rate fluctuations, exchange rates may still fluctuate owing to changes in market fundamentals. A government could, in principle, try to fix the exchange rate in this case, but doing so would entail a loss in sovereignty over the conduct of domestic monetary policy. In any case, the imposition of legal restrictions on foreign currency holdings is not without cost, since they hamper the conduct of international trade (e.g., individuals are forced to make currency conversions that they would otherwise prefer not to make).

A more dramatic policy may entail a return to the past, where governments issued monetary instruments that were backed by gold. In general, governments might also issue money that is backed by other real assets (like domestic real estate). Under this scenario, government money would presumably trade much like any private security. The value of government money would depend on both the value of the underlying asset backing the money and the government's willingness and/or ability to make good on its promises. The relative price of national monies would then depend on the market's perception of the relative credibility of competing governments. Governments typically do not like to issue money backed in this manner, since it restricts their ability to extract seigniorage revenue and otherwise conduct monetary policy in a 'flexible' manner.

But perhaps the ultimate solution may entail removing the government monopoly on paper money. By many accounts, historical episodes in which private banks issued (fully-backed) money appeared to work reasonably well (e.g., the so-called U.S. 'free-banking' era of 1836-63). Despite problems of counterfeiting (which are obviously present with government paper as well) and despite the coexistence of hundreds of different bank monies, these monies generally traded more often than not at relatively stable fixed exchange rates.[7] Such a regime, however, severely limits the ability of governments to collect seigniorage revenue. It is no coincidence that the U.S. 'free-banking' system was legislated out of existence during a period of severe fiscal crisis (the U.S. civil war).

So there you have it. Given the political landscape, it appears that no monetary system is perfect. Each system entails a particular set of costs and benefits that continue to be debated to this day.

11.5 References

1. Champ, Bruce and Scott Freeman (2001). *Modeling Monetary Economies*, 2nd Edition, Cambridge University Press, Cambridge, U.K.

2. King, Robert, Neil Wallace, and Warren E. Weber (1992). "Nonfundamental Uncertainty and Exchange Rates," *Journal of International Economics*, 32: 83–108.

3. Manueli, Rodolfo E. and James Peck (1990). "Exchange Rate Volatility in an Equilibrium Asset Pricing Model," *International Economic Review*, 31(3): 559–574.

4. Roubini, Nouriel, Giancarlo Corsetti, and Paolo A. Pesenti (1998). "Paper Tigers? A Model of the Asian Crisis," Manuscript.

5. Spiegel, Mark (2002). "Argentina's Currency Crisis: Lessons for Asia," www.frbsf.org/ publications/ economics/ letter/ 2002/ el2002-25.html

[7] The episodes in which some banknotes traded at heavy discount were often directly related to state-specific fiscal crises and legal restrictions that forced state banks to hold large quantities of state bonds.

11.A Nominal Exchange Rate Indeterminacy and Sunspots

The indeterminacy of nominal exchange rates opens the door for multiple rational expectations equilibria (self-fulfilling prophesies). This appendix formalizes the restrictions placed on exchange rate behavior implied by our model; see also Manuelli and Peck (1990).

Consider two countries a and b as described in the text. Individuals have preferences given by $E_t u(c_2(t+1))$, where E_t denotes an expectations operator. Note that there is no fundamental risk in this economy. But there may be nonfundamental risk owing to fluctuations in the exchange rate.

Let R^i_{t+1} denote the (gross) rate of return on currency $i = a, b$ (i.e., the inverse of the inflation rate). Let q^i_t denote the real money balances held of currency i by an individual. Then an individual faces the following set of budget constraints:

$$q^a_t + q^b_t = y;$$
$$c_2(t+1) = R^a_{t+1} q^a_t + R^b_{t+1} q^b_t.$$

Substitute these constraints into the utility function. The individual's choice problem can then be stated as:

$$\max_{q^a_t} E_t u\left(\left[R^a_{t+1} - R^b_{t+1}\right] q^a_t + R^b_{t+1} y\right).$$

A necessary (and sufficient) condition for an optimal choice of q^a_t is given by:

$$E_t \left[R^a_{t+1} - R^b_{t+1}\right] u'(c_2(t+1)) = 0.$$

Since the marginal utility of consumption is assumed to be positive, the condition above implies:

$$E_t \left[R^a_{t+1} - R^b_{t+1}\right] = 0.$$

Let $e_t = p^a_t / p^b_t$ denote the nominal exchange rate. Then $R^a_{t+1} = p^a_t / p^a_{t+1}$ and $R^b_{t+1} = p^b_t / p^b_{t+1}$. Hence, $(e_{t+1}/e_t) = R^b_{t+1}/R^a_{t+1}$ or $R^b_{t+1} = (e_{t+1}/e_t) R^a_{t+1}$. Substitute this latter condition into the previous equation, so that:

$$E_t \left[1 - \left(\frac{e_{t+1}}{e_t}\right)\right] R^a_{t+1} = 0.$$

Since $R^a_{t+1} > 0$, this condition implies:

$$E_t e_{t+1} = e_t.$$

This condition exhausts the restrictions placed on exchange rate behavior by the theory. What the condition tells us is that the nominal exchange rate can be stochastic, but that the stochastic process must follow a *Martingale*. In other words, the best forecast for the future exchange rate e_{t+1} is given by the current (indeterminate) exchange rate e_t. A special case is given by a deterministic exchange rate; i.e., $e_{t+1} = e_t$. But many other outcomes are possible so that the equilibrium exchange rate may fluctuate even in the absence of any intrinsic uncertainty (i.e., no fundamental risk).

Note that, in equilibrium, q_t^a and q_t^b represent the *average* real money holdings of individuals. These demands can fluctuate with the appearance of 'sunspots.' For example, if individuals (for some unexplained reason) feel like 'dumping' currency a, then q_t^a will fall. Of course, since $q_t^a + q_t^b = y$, such behavior implies a corresponding increase in the demand for currency b. If individuals are risk-averse (i.e., if $u'' < 0$), then individuals would want to hedge themselves against any risk induced by sunspot movements in the exchange rate. One way to do this is for all individuals to hold the average quantities q_t^a and q_t^b in their portfolios. In this way, any depreciation in currency a is exactly offset by an appreciation in currency b. However, if individuals find it costly to hedge in this manner, then sunspot uncertainty will induce 'unnecessary' variability in individual consumptions, leading to a reduction in economic welfare.

11.B International Currency Traders

In the model of exchange rate indeterminacy developed in the text, it was assumed that all individuals view different fiat currencies as perfect substitutes. In fact, all that is required for the indeterminacy result is that *some* group of individuals view fiat currencies as perfect substitutes. In reality, this some group of individuals can be thought of as large multinational firms that readily hold assets denominated in either USD, Euros, or Yen (for example). In this case, indeterminacy will prevail even if the domestic residents of (say) the United States and Japan each prefer (or are forced by legal restriction) to hold assets denominated in their national currency.

This idea has been formalized by King, Wallace and Weber (1992). To see how this might work, consider extending our model to include three types of individuals A, B, and C, with each type consisting of a fixed population N. Think of type A individuals as Americans living in the U.S. (country a) and type B individuals as Japanese living in Japan (country b). Type C individuals are 'international' citizens living in some other location (perhaps a remote island in the Bahamas).

Assume that foreign currency controls force domestic residents to hold domestic currency only. International citizens, however, are free to hold either currency. Let q_t^i denote the real money balances held by international citizens in the form of $i = a, b$ currency. Note that $q^a + q^b = y$. Then the money-market clearing conditions are given by:

$$M^a = p_t^a N(y + q_t^a);$$
$$M^b = p_t^b N(y + q_t^b).$$

The nominal exchange rate in this case is given by:

$$e_t = \frac{p_t^a}{p_t^b} = \frac{M^a}{M^b}\left(\frac{y + q_t^b}{y + q_t^a}\right) = \frac{M^a}{M^b}\left(\frac{2y - q_t^a}{y + q_t^a}\right).$$

This condition constitutes one equation in the two unknowns: e_t and q_t^a. Hence, the exchange rate is indeterminate and may therefore fluctuate solely on the 'whim' of international currency traders (i.e., via their choice of q_t^a). If international currency traders are well-hedged (or if they are risk-neutral), exchange rate volatility does not matter to them. But any exchange rate volatility will be welfare-reducing for the domestic residents of countries a and b.

11.C The Asian Financial Crisis

Perhaps you've heard of the so-called *Asian Tigers*. This term was originally applied to the economies of Hong Kong, South Korea, Singapore and Taiwan, all of which displayed dramatic rates of economic growth from the early 1960s to the 1990s. In the 1990s, other southeast Asian economies began to grow very rapidly as well; in particular, Thailand, Malaysia, Indonesia and the Philippines. These 'emerging markets' were subsequently added to the list of Asian Tiger economies. In 1997, this impressive growth performance came to a sudden end in what has subsequently been called the Asian Financial Crisis. What was this all about?

Throughout the early 1990s, many small southeast Asian economies attracted huge amounts of foreign capital, leading to huge net capital inflows or, equivalently, to huge current account deficits. In other words, these Asian economies were borrowing resources from the rest of the world. Most of these resources were used to finance domestic capital expenditure. As we learned in Chapters 4 and 6, a growing current account deficit may signal the strength of an economy's future prospects. Foreign investors were forecasting high future returns on the capital being constructed in this part of the world. This 'optimism' is what fuelled much of the growth domestic capital expenditure, capital inflows, and general growth in these economies. Evidently, this optimistic outlook turned out (after the fact) to be misplaced.

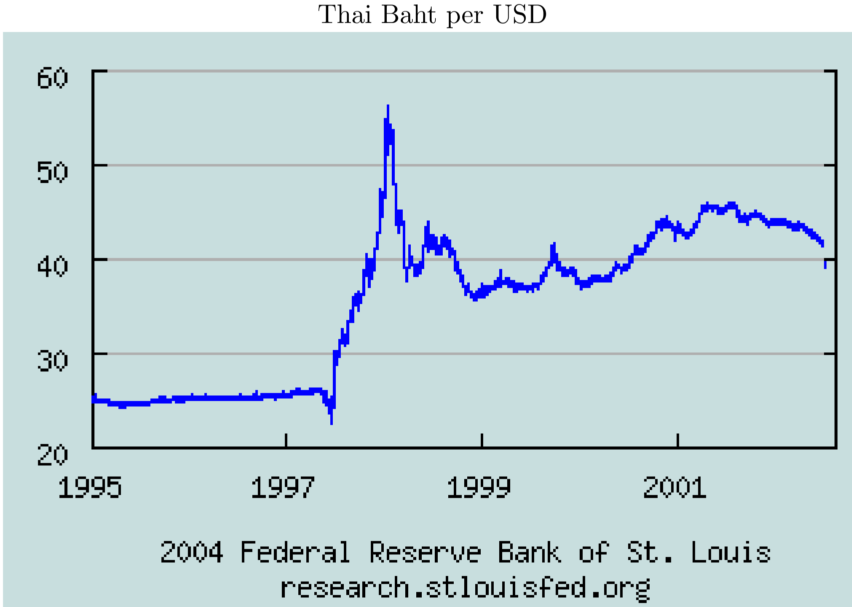
What went wrong? One possible is that nothing went 'wrong' necessarily. After all, rational forecasts can (and often do) turn out to be incorrect (after the fact). Perhaps what happened was a growing realization among foreign investors that the high returns they were expecting were not likely to be realized. Investors who realized this early on pulled out (liquidating their foreign asset holdings). As this realization spread throughout the world, the initial trickle in capital outflows exploded into a flood. Things like this happen in the process of economic development.

Of course, there are those who claim that the 'optimism' displayed by the parties involved was 'excessive' or 'speculative;' and that these types of booms and crashes are what one should expect from a free market. There is another view, however, that directs the blame toward domestic government policies; e.g., see (Roubini, Corsetti and Pesenti, 1998). For example, if a government stands ready to bailout domestic losers (bad capital projects), then 'overinvestment' may be the result as private investors natural downplay the downside risk in any capital investment. To the extent that foreign creditors are willing to lend to domestic agents against future bail-out revenue from the government, unprofitable projects and cash shortfalls are refinanced through external borrowing. While public deficits need not be high before a crisis, the eventual refusal of foreign creditors to refinance the country's cumulative losses forces the government to step in and guarantee the outstanding stock of external liabilities. To satisfy solvency, the government must then undertake appropriate domestic fiscal reforms, possibly involving recourse to seigniorage revenues. Expectations of inflationary financing thus cause a collapse of the currency and anticipate the event of a financial crisis.

There is also evidence that government corruption may have played a significant role. For example, a domestic government may borrow money from foreigners with the stated intent of constructing domestic capital infrastructure. But a significant fraction of these resources may simply be 'consumed' by government officials (and their friends). For example, in 2001 Prime Minister Thaksin (of Thailand) was indicted for concealing huge assets when he was Deputy Prime Minister in 1997. Evidently, Mr. Thaksin did not dispute the charge. Instead, he said that the tax rules and regulations were 'confusing' and that he made an 'honest mistake' in concealing millions of dollars assets, manipulating stocks and evading taxes.[8] As we saw during the recent *Enron* scandal, the market reacts quickly and ruthlessly when it gets a whiff of financial shenanigans.[9]

The Asian crisis began in 1997 with a huge speculative attack on the Thai currency (called the *Baht*). Prior to 1997, the Thai government had unilaterally pegged their currency at around 25 Baht per USD. Many commentators have blamed this speculative attack for precipitating the Asian crisis. A more plausible explanation, however, is that the speculative attack was more of a *symptom* than a cause of the crisis (which was more deeply rooted in the nature of government policy).

Thai Baht per USD



[8]See: www.pacom.mil/ publications/ apeu02/ 24Thailand11f.doc.

[9]For more on the Enron fiasco, see: www.washingtonpost.com/ wp-dyn/ business/ specials/ energy/ enron/

In any case, financial crisis or not, our theory suggests that the Thai government could have maintained its fixed exchange rate policy and prevented a speculative attack on its currency if it had either: (1) maintained sufficient USD reserves; or (2) been willing to tax its citizens to raise the necessary USD reserves. Evidently, as the Thai economy showed signs of weakening in 1997, currency speculators believed that neither of these conditions held (and in fact, they did not).

Would the crisis in Thailand have been averted if the government had maintained a stable exchange rate? It is highly doubtful that this would have been the case. If the crisis was indeed rooted in the fact that many bad investments were made (the result of either bad decisions or corruption), then the contraction in capital spending (and the corresponding capital outflows) would have occurred whether the exchange rate was fixed or not. But this is just one man's opinion.

Chapter 12

Money, Capital and Banking

12.1 Introduction

An almost universal feature of most economies is the coexistence of 'base' and 'broad' money instruments. In modern economies, the monetary base consists of small denomination government paper notes, while broad money consists of electronic book-entry credits created by the banking system redeemable in base money. In earlier historical episodes, the monetary base often consisted of specie (gold or silver coins) with broad money consisting of privately-issued banknotes redeemable in base money. In this chapter, we consider a set of simple models designed to help us think about the determinants of such monetary arrangements.

12.2 A Model with Money and Capital

Consider a model economy consisting of overlapping generations of individuals that live for two periods. The population of young individuals at date $t \geq 1$ is given by N_t and the population grows at some exogenous rate $n > 0$; i.e., $N_t = nN_{t-1}$. There is also a population N_0 of individuals we call the 'initial old' (who live for one period only). To simplify, we focus on stationary allocations and assume that individuals care only about consumption when old; i.e., $U(c_1, c_2) = c_2$.

As before, assume that each young person has an endowment $y > 0$. Unlike our earlier OLG model, however, assume that output can be stored over time in the form of capital. In particular, k units of capital expenditure at date t

yields $zf(k)$ units of output at date $t+1$, where f is increasing and strictly concave in k. Assume (for simplicity) that capital depreciates fully after its use in production. Note that $zf'(k)$ represents the marginal product of (future) capital. From Chapter 6, we know that $zf'(k)$ will (in equilibrium) be equal to the (gross) real rate of interest.

It will be useful to consider first how this economy functions without money. The equilibrium, in this case, is simple to describe. In particular, since young individuals care only for future consumption, they will save their entire endowment in the form of capital spending; i.e., $k = y$. In this way, the young end up consuming $c_2 = zf(y)$. The economy's GDP at any given date t is given by:

$$Y_t = N_t y + N_{t-1} zf(y).$$

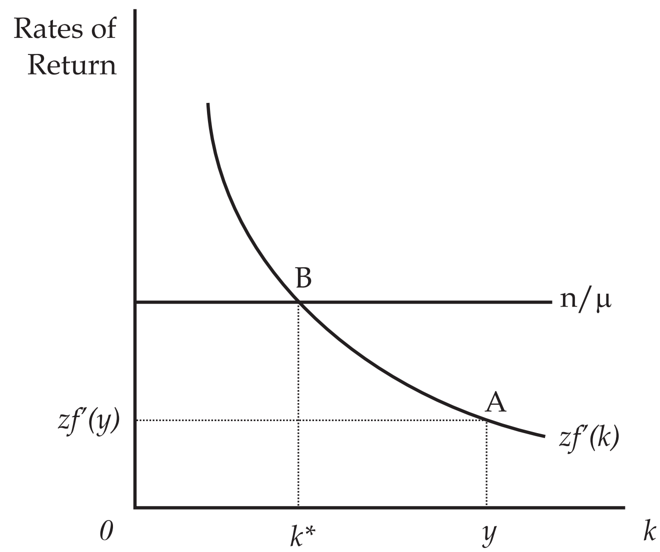
Let us now introduce money into this economy. Assume that there is a supply of fiat money M_t that is initially (i.e., as of date 1) in the hands of the initial old. The money supply grows at an exogenous rate $\mu \geq 1$ so that $M_t = \mu M_{t-1}$. Assume that new money is used to finance government purchases. Let Π denote the (gross) rate of inflation, so that Π^{-1} represents the (gross) real return on fiat money. From Chapter 10, we know that in an equilibrium where fiat money is valued, its rate of equilibrium rate of return will be given by $\Pi^{-1} = n/\mu$.

A young individual now faces a portfolio choice problem. That is, there are now two ways in which to save for future consumption: money and capital. Let q denote the real money balances acquired by a young individual (from the existing old generation). The portfolio choice is restricted to satisfy $q + k = y$. A simple no-arbitrage condition implies that both money and capital must earn the same rate of return (if both assets are to be willingly held). Therefore, the equilibrium level of capital spending must satisfy:

$$zf'(k^*) = \frac{n}{\mu}. \tag{12.1}$$

This equilibrium is displayed in Figure 12.1.

FIGURE 12.1
Coexistence of Money and Capital



Point A in Figure 12.1 displays the equilibrium without money. In this scenario, the level of capital investment is 'high' so that the return to capital is 'low' (recall that we are assuming a diminishing marginal product of capital). If $n/\mu > zf'(y)$, then young individuals are willing to divert some of their savings away from capital into money. As capital spending declines, the marginal product of capital rises until the return on capital is equated to the return on money (point B).

- **Exercise 12.1.** Consider Figure 12.1. Explain why the welfare of all individuals (including the initial old) is higher at point B relative to point A. Explain why this is so despite the fact that real GDP is lower at an equilibrium associated with point B.

- **Exercise 12.2**. Use the model developed above to explain the likely economic consequences of a decline in z. Hint: recall that z here has the interpretation of being the current period *forecast* of future productivity. In particular, emphasize the likely impact of this 'bad news' shock on capital spending, the demand for money, and future GDP.

12.2.1 The Tobin Effect

Equation (12.1) suggests that while monetary policy may be neutral, it is unlikely to be superneutral.[1] Consider, for example, an exogenous (permanent) increase in the money growth rate μ. In this model, such a policy has the effect of increasing the expected and actual rate of inflation. Since the real rate of return on money is inversely related to the inflation rate, the expected return on money falls. If the return on money falls, equation (12.1) suggests that individuals will substitute out of money and into capital (i.e., rebalance their wealth portfolios) to a point at which all assets again earn the same rate of return (see also Figure 12.1). Since the real GDP in this economy is given by:

$$Y_t = N_t y + N_{t-1} z f(k_{t-1}),$$

it follows that such a 'loosening' of monetary policy has the effect of expanding the future GDP (as the expansion in current capital spending adds to the future capital stock). The substitution of private capital for fiat money in reaction to an anticipated inflation is called the *Tobin effect* (Tobin, 1965).

The Tobin effect appears to present policymakers with a policy tool that may be used to 'stimulate' the economy during a period of economic recession or stagnation. In fact, some economists have advocated such a policy for Japan, which struggled throughout the 1990s with low economic growth and low inflation (and even deflation). In a deflationary environment, the rate of return on money is high, making capital investment relatively unattractive. For this reason, many people view deflation as undesirable leading them to advocate a policy geared toward increasing the inflation rate.

The logic underpinning the Tobin effect is compelling enough, but a few caveats are in order. First of all, it is important not to confuse GDP with economic welfare. An inflation-induced expansion of the capital stock (and GDP) in our model unambiguously *reduces* the welfare of all individuals. In particular, a very high inflation would move the economy from point B to point A in Figure 12.1 (see exercise 12.1). One can, in fact, show that the 'optimal' money growth rate in this model is zero (achieved by setting $\mu = 1$), which (for $n > 1$) implies that deflation is optimal.[2]

There are also some practical difficulties associated with exploiting the Tobin effect in reality. The first of these is that the stock of fiat money is typically tiny relative to the size of the stock of capital. Thus, while one may accept the logic of the Tobin effect, one may legitimately question whether the effect can be quantitatively important (probably not). The second problem is that while our model features a clear link between money growth and inflation expectations,

[1] That is, note that capital spending (and other real variables) does not depend on the money stock M_t; but does depend on the money growth rate μ.

[2] Few people appear to take seriously the notion that deflation is optimal. This view has largely been shaped by the experience of the Great Depression. However, there are many (earlier) historical episodes in which booms were associated with deflations. These episodes appear to have been erased from society's collective memory bank.

in reality this link appears to be not so tight. In Japan, for example, inflation expectations appear to remain low even today despite several years of rapidly growing government money and debt.[3]

12.3 Banking

In the model developed above, fiat money and physical capital are viewed as perfect substitutes; i.e., each asset serves the same purpose and earns the same rate of return. In some sense, physical capital in this model is a type of 'private money' that competes with government fiat money. In the model, this private money can be thought of as privately-issued notes representing claims against capital (e.g., corporate bonds). In equilibrium, individuals are indifferent between getting paid in government or private paper.

In reality, most privately-issued debt instruments are not widely used as a means of payment. An important exception to this general rule are the liabilities issued by banks; sometimes called *demand-deposit liabilities*. If you have a bank account, then most your money (not wealth) is likely held in the form of this 'bankmoney' (electronic credits in a checkable account). Every time you use your debit card or write a check, you are in effect making a payment with bankmoney (i.e., the banking system simply debits your account electronically and makes corresponding credit entries to the accounts of merchants–no paper money is involved in such transactions).

But we do not use bankmoney for all of our transactions. For some transactions, we find it convenient to use government money (cash). Few of us actually get paid in cash; our employers pay us in bankmoney (i.e., by writing a check or making a direct transfer to our bank account). When we need cash, we can visit a bank or ATM and make a withdrawal. Note that our bankmoney constitutes a demandable liability of the bank. That is, we can visit an ATM at any time and demand the redemption of our bankmoney for cash (at par value).

Several questions may be popping to mind here. How are we to understand the coexistence of fiat money and private money? Why is private money almost always made redeemable on demand for cash (i.e., either government fiat or historically in the form of specie)? Are banks free to 'print' all the bankmoney they want? What is the function of banks–what do they do? Are they simply repositories for cash, or do they serve some other function? What might happen if all depositors wanted to withdraw cash from the banking system at the same time? To help organize our thinking on these matters, let us develop a simple model.

[3]One possible explanation for this is that the Japanese are expecting the government to pay back the debt out of direct tax revenue instead of seigniorage.

12.3.1 A Simple Model

The formal model associated with this section can be found in Andolfatto (200*). In what follows here, we will make do with an informal description. The economy is similar to that described above. Imagine, however, that individuals populate two 'spatially separated' locations, labelled A and B. Each location is identical in terms of population, technology and endowments. Assume that individuals living in a location do not accept private liabilities issued by agents living in the 'foreign' location. There are several ways in which to interpret this restriction. One interpretation is that local residents do not trust 'foreign' paper, perhaps because it is difficult to detect counterfeits (unlike locally issued money or government cash). Another interpretation is that the two locations are not connected electronically, so that debit card transactions are not feasible for 'tourists' visiting the foreign location.

Young individuals have an endowment y and have access to an investment technology that takes k units of output today and returns zk units of output in the future; i.e., assume constant returns to scale so that $f(k) = k$. In this case, the (gross) return to capital spending is given by z. Assume that $z > n$ and that $\mu = 1$ (constant supply of fiat money). In this case, capital dominates money in rate of return (as is typically the case in reality). But if capital (private money) dominates fiat money, why is the latter held at all? In order to be valued, fiat money cannot be a perfect substitute for capital–it needs to fulfill some other purpose.

This other purpose is generated in the following way. Young individuals must make a portfolio decision (money versus capital). Imagine that after this decision is made, a fraction $0 < \pi < 1$ of young individuals at each location are exogenously transported to the foreign location. Assume that they cannot take their capital with them. They could try to take paper notes representing claims to the capital they left behind, but by assumption, such claims are not 'recognized' in the 'foreign' location. How then are these 'visiting' individuals to purchase the consumption they desire when old? One way would be to use a payment instrument that is widely recognizable, like government fiat or gold. In this sense then, it is useful to think of π as an the probability of experiencing an individual 'liquidity' shock (i.e., of encountering a situation where merchants will only accept cash).

Since young individuals are uncertain about whether they will need cash in a future transaction, they will want to hedge their bets (if they are risk-averse) by holding some capital and some cash in their portfolio. Hedging in this way is better than not hedging at all, but it is still inefficient. Since there is no aggregate uncertainty, it would make sense for young individuals to pool their risk. One way of doing this is through a bank that operates in the following way.

Imagine that young individuals 'deposit' their endowment (real labor earnings) with a bank. In return, the bank offers each depositor an interest-bearing account that is redeemable (on demand) for government cash. If individuals do not make an early withdrawal, they earn a real interest rate equal to z (and consume locally). If they make an early withdrawal (because the experience the liquidity shock), they earn a real return equal to $n < z$ (and consume in the foreign location).

A competitive bank would structure its balance sheet in the following way. It would take the deposits made to the bank and use some of the deposits to acquire cash (from the existing old agents). How much cash a bank should hold in reserve depends on how much cash is forecasted to be withdrawn. In our model, there is no aggregate uncertainty so that the bank can easily forecast that a known fraction π of its depositors will want to make an early withdrawal. Any resources that are not used up in the form of cash reserves can then be used to finance capital spending (e.g., making loans to entrepreneurs or acquiring the capital directly). The bank can then use the return on its loan portfolio to pay the interest its own liabilities (i.e., the interest on its deposit accounts).

A few observations are in order. First, note that the balance sheet of our model bank resembles the balance sheet of real banks. That is, the asset side consists of loans and cash reserves. The liability side consists of demandable debt. This demandable debt earns a higher return than government cash and is used widely as a form of payment. But not all places accept bankmoney (checks or debit cards). For this reason, banks stock their ATMs with government cash and allow their depositors to withdraw this cash on demand.

Second, note that banks are not free to create unbacked money (unlike the government). A well-run bank will be 'well-capitalized' in the sense that it should have assets of sufficient value to back up its outstanding liabilities. Note, however, this does not mean a bank will hold all of its assets in the form of government cash (which would correspond to a 100% reserve requirement). By holding less than 100% reserves, banks are able to facilitate the financing of productive capital projects (that might not otherwise get financed). Of course, banks can be subject to fraudulent activity; for example, if a loans officer grants a loan to a friend (who has no intention of repaying) and tries to pass it off as an 'investment.' But fraud is illegal and is, in any case, not specific to the banking industry.

Finally, there is an issue of whether banks are in some sense 'fragile' financial structures that may be subject to 'bank runs' (where everyone rushes to the bank to withdraw cash). The reason people worry about this is owing to the structure of a bank's balance sheet. In particular, its assets are relatively 'long-term' and 'illiquid,' while its liabilities constitute a form of very short-term debt (i.e., demandable debt). Of course, I argued above that this balance sheet structure is no accident; i.e., this is the business of banking. But the question remains as to whether such a structure is susceptible to collapse that is generated, for example, by a self-fulfilling prophesy. What if, for example, everyone suddenly perceives a bank to be insolvent. Then they will 'run' to the bank to withdraw their cash. If everyone runs, there will not be enough cash to satisfy depositors.

The bank will then be forced to liquidate its assets quickly at firesale prices, possibly rendering the bank insolvent and confirming the initial expectation.

Certainly, there have been historical episodes that seem to fit the description above. It is not clear, however, whether such episodes were the product of banking *per se* or of government restrictions on bank behavior. For example, branch banking was for many years prohibited in the United States but not in Canada, leading to thousands of smaller banks in the U.S. and only a few larger banks in Canada. It is well-known that many banks failed in the U.S. during the Great Depression whereas none did in Canada. In another example, many of the state banks of the so-called U.S. 'free-banking' era were forced to hold as assets state bonds of questionable quality. When state governments fell into fiscal crisis, banks that held government bonds naturally did too. Both of these examples are often cited as evidence of banking 'instability.' However, in both cases, legal restrictions were in place that prevented banks from creating a well-diversified asset portfolio.

12.3.2 Interpreting Money Supply Fluctuations

As I have remarked before, much of an economy's money supply is in the form of demand deposit liabilities created by private banks. These liabilities are convertible on demand (and at par) for government cash. The fact that bankmoney is convertible at par with government cash effectively fixes the nominal exchange rate between bankmoney and fiat money. One popular definition of an economy's total supply of money is therefore given by the sum of government cash (in circulation) and the demandable liabilities created by chartered banks. This definition of money is called $M1$. Broader measures of money may include other types of private securities, including savings accounts, term deposits, money market mutual fund balances, and so on.

Most of the fluctuations in an economy's money supply stems from fluctuations in bankmoney, or the so-called 'money-multiplier.' The money multiplier is defined as the total money supply divided by (essentially) the monetary base (i.e., government cash). Fluctuations in the money multiplier (or broad money) appear to be closely linked with fluctuations in real output. Changes in the current money supply appear to be positively correlated with future changes in *real* GDP. Why does the money supply vary in this manner? Does the observed money-output correlation imply that a monetary authority stimulate the economy with a surprise injection of cash?

To help address these questions, let us consider the model developed above (and stated formally in the appendix). Assume that the stock of government money remains fixed at M. Let (q^*, k^*) denote the bank's (or banking system's) optimal portfolio choice. Note that this choice will depend on the model's parameters. The parameter I wish to emphasize here is z; i.e., the current forecast of the future return to capital investment. Under a mild restriction on preferences, the model implies that k^* is an increasing function of z. In other words, capital spending today responds positively to 'news' of a higher expected return on capital (note that z also corresponds to the real rate of interest in this economy). Since $q + k = y$, it follows that the demand for real money balances is a decreasing function of z. In other words, news of a higher z leads market participants (including banks) to substitute out of cash and into capital (banks would accomplish this by extending more loans to finance capital spending).

Now imagine that the economy is continually hit by shocks that take the form of changes in z. What are the economic implications of such disturbances and, in particular, how do they affect the money supply?

Let us begin with the money-market clearing condition:

$$\frac{M_t}{p_t} = N_t q_t.$$

For simplicity, let us hold the supply of fiat money and the population fixed. Then the equilibrium price-level will satisfy:

$$p_t^* = \frac{M}{N q_t^*}. \tag{12.2}$$

- **Exercise 12.3**. Imagine that the economy receives 'news' that leads individuals to revise downward their forecast of z. What effect is such a shock likely to have on the real interest rate and the price-level? Explain.

Now recall that the total money supply ($M1_t$) is the sum cash and bankmoney. The real value of a bank's physical capital is given by Nk_t^* (assuming that it has N depositors). The nominal value of this bankmoney is given by $p_t^* N k_t^*$, so that:

$$M1_t = M + p_t^* N k_t^*.$$

Substituting equation (12.2) into this latter expression, one can derive:

$$\begin{aligned} M1_t &= M + \frac{M}{N q_t^*} N k_t^*; \\ &= \left[1 + \frac{k_t^*}{q_t^*} \right] M. \end{aligned} \tag{12.3}$$

The term in the square brackets above is the money-multiplier $M1/M$. Our model suggests that the money multiplier is positively related to the 'deposit-to-currency' ratio k/q.

Now consider the empirical observation that changes in the current money supply $M1_t$ appear to be positively correlated with future changes in real GDP. We can use the model developed here to interpret such a correlation. Suppose, for example, that the economy is subject to 'news' shocks that lead people to constantly revise their forecasts of z (the future return to current capital spending). Consider, for example, a sudden increase in z. From our earlier discussion, such a shock should lead individuals (or banks acting on their behalf) to substitute out of cash and into capital investment. In other words, k_t^* should increase and q_t^* should decrease. From equation (12.3), we see this reaction will lead to an increase in the money multiplier and hence to an increase in $M1_t$ (even if M remains constant). Furthermore, the increase in current period capital spending will translate (together with the increase in z, if it materializes) into higher levels of future GDP.

Notice that while our model can replicate the observed money-output correlation, the model suggests that it would be wrong to interpret the current increase in $M1_t$ as having 'caused' the future increase in real GDP. In fact, the direction of causality is reversed here; i.e., it is the increase in (forecasted) real GDP that 'causes' an increase in the current money supply. This is an example of what economists call 'reverse causation,' and serves as a useful warning for us not to assume a direction of causality simply on the basis of an observed correlation (e.g., we would not, for example, assert that Christmas shopping 'causes' Christmas, even though though early-December Christmas shopping is positively correlated with the future arrival of Christmas).

12.4 Summary

To be written (chapter is incomplete).

12.5 References

1. Andolfatto, David (2003). "Taking Intermediation Seriously: A Comment," Journal of Money, Credit and Banking, 35(6-2): 1359–1366.

2. Tobin, James (1965). "Money and Economic Growth," *Econometrica*, 33: 671–84.

Part III

Economic Growth and Development

Chapter 13

Early Economic Development

13.1 Introduction

The questions of why some economies grow while others do not, or why economies grow at all, are perhaps the most fascinating (and unresolved) issues in the science of economics. These questions are related, but distinct. The former question asks why less-developed economies simply do not imitate their well-developed counterparts. The latter question asks what triggers economic development in even relatively advanced economies. Given that our current living standards depend on past growth rates and that our future living standards (or those of our children) will depend on current and future growth rates, the question of growth and development is of primary importance.

Most people would agree that the high living standards that we enjoy today (relative to historical levels and to less developed contemporary economies) is directly attributable to the advanced state of technology, which we can define broadly to include basic scientific and engineering knowledge, human capital, organizational design, and so on. Understanding this basic fact, however, is not very helpful since it does not explain how we came to acquire this advanced technology (and why some less developed economies simply do not imitate us). Furthermore, it appears evident that there must be more to the story of rising per capita living standards than just technological progress. As I will describe in this chapter, technological progress has been with us since the beginning of recorded history. And yet, up until the Industrial Revolution (c. 1750), growth in per capita living standards appears to have been modest at best. It appears that technological advances in the Pre-Industrial Revolution era manifested themselves primarily in the form of population growth (i.e., growth in GDP rather than per capita GDP).

In this chapter, I will take some time to outline the evidence pertaining to technological advancements prior to the Industrial Revolution. I will then present a theory of growth (due to Thomas Malthus). Like many so-called growth models, however, the theory does not explain the source of technological advancement. But the model does offer an explanation as to why technological progress (if it occurs at all) manifests itself as population growth, instead of growth in per capita living standards.

13.2 Technological Developments

This section is not meant to be an exhaustive survey of technological progress throughout the ages. But a sufficient number of examples are provided that should convince you of the remarkable technological advancements that have occurred throughout recorded history.[1] As is the case today (but to a lesser extent), economies varied in their use of available technologies. The diffusion (implementation) of new technologies (and ideas in general) occurred slowly (if at all). Economies that employed the best technology often grew to be world powers, dominating those around them. It is also interesting to note that the centers of power (wealth and knowledge) shifted many times over the course of history.

13.2.1 Classical Antiquity (500 B.C. - 500 A.D.)

The classical civilizations of Europe (Greek, Hellenistic, and Roman) were in many ways more scientifically advanced than anything Europe had to offer up to the renaissance period (c. 1450 A.D.). Important achievements of classical technology included: coinage, alphabetization, stenography and geometry. In addition to their achievements in literature, advances were made in mathematics, medicine, science, architecture, construction, and political organization. However, in other areas, there was (from our perspective today) a distinct lack of progress in many important areas, such as the development of mechanical devices, chemistry, and farming.

The technological progress in the classical world (especially in Roman times) was typically geared to the public, rather than the private, sector. The Rome of 100 A.D. had better paved streets, sewage, water supply, and fire protection that the capitals of Europe in 1800 A.D. Most agriculture, manufacturing, and services were carried out by the private sector, and achievements there were slow. The lack of progress in developing mechanical devices for commercial uses is particularly interesting, since a great achievement of the classical world was in recognizing fully the important elements of machinery, such as the lever, the wedge and screw, the ratchet, the pulley, the gear and the cam. Apparently,

[1] These examples are taken from Mokyr (1990).

these insights were applied mostly to war machines and clever gadgets that were admired for their own sake but rarely put to useful purposes.

Many of the classical ideas lay dormant for centuries. For example, Hero of Alexandria (1st century A.D.) developed a working steam engine used to open temple doors; a coin-operated vending machine (for holy water in the temple). Similarly, Ctesbius (3rd century A.D.), who has been called by some the Edison of Alexandria, reportedly invented the hydraulic organ, metal springs, the water clock, and the force pump. An important question is why so little of this potential was realized and translated into economic progress. Many inventions that could have led to major economic changes were underdeveloped, forgotten or lost. This is especially puzzling in light of the fact that classical civilizations were relatively literate and mobile.

13.2.2 The Middle Ages (500 A.D. - 1450 A.D.)

Early medieval Europe (sometimes referred to as the *Dark Age*) managed to break through a number of technological barriers that held the Romans back. This is all the more remarkable when one recognizes that many of the ingredients that are commonly thought of as essential to technological progress were absent. Especially during the Dark Ages, the economic and cultural climate in Europe was primitive compared to the Classical period. Literacy had become rare and the upper classes devoted themselves primarily to warfare. Commerce and communications declined to almost nothing. The great infrastructure of the Roman empire was left to depreciate. Law enforcement and the security of life and property became more precarious than ever. And yet toward the end of the Dark Ages, in the eighth and ninth centuries, European society began to show the first signs of what was soon to become a torrent of technological creativity. In contrast to the Classical age, these new technological developments took the form of practical tools and ideas that reduced the daily toil and increased the material comfort (however marginally) of the masses.

In terms of their direct contribution to aggregate output, changes in agricultural technology were of primary importance as the bulk of the population was engaged in farming. Some of the key advances included: the introduction of the heavy plow, the wheelbarrow (both invented in China, c. 300-100 B.C.) and the three-field system; the extensive use of water power; and the modern horse collar (and horseshoe). The widespread application of these technologies allowed medieval Europe to build an economy based on nonhuman power, in contrast to the widespread use of slaves in Classical economies. Other important inventions included the weight-driven mechanical clock, which has been referred to some as the key machine of the modern industrial age, and the printing press, which allowed books to be published on an unprecedented scale.

Medieval Western technology drew from three sources: classical antiquity, contemporary Islamic and Asian societies, and its own original creativity. The second source of ideas is of particular interest, since it indicates a willingness by Europeans to shamelessly adopt the ideas of other cultures.

From about 700 A.D. - 1200 A.D., the cultural and technological center of gravity of Europe remained to a large extent in the Islamic societies of Spain, Northern Africa and the Middle East. The culture and technology of Islam constituted a synthesis of Hellenistic and Roman elements, together with ideas from central Asia, India, Africa and China. Early Islamic society collected, compiled, and catalogued knowledge avidly. Both rich and poor alike apparently travelled extensively and learned eagerly from other cultures. It was a society literate beyond Europe's wildest dreams. During this time, the Moslems knew more about the different parts of the known world than any other civilization.

Important Moslem contributions to technology in this era include the development of the lateen sail, tidal mills, and windmills. They were also responsible for the introduction of paper (invented in China, c. 100 A.D.) into the Middle East and Europe (the first paper factory was set up in Baghdad, c. 793 A.D.). Other notable advances occurred in textile production, agriculture, and chemical technology. The great strength of the early Islamic world, however, lay in its ability and willingness to adopt existing technologies from foreign sources.

From about 1200 A.D. on, however, both innovative and imitative activity in the Moslem world slowed down, for reasons that are still not very clear. Some people blame the destruction of the eastern Islamic world by Mongol invasions. Others point to political and religious factors. In any case, by about this time, the economies of western Europe had absorbed most of what the Moslem world had to offer.

The truly great technological powerhouse of the ancient and medieval world was China, which maintained its technological supremacy until about 1400 A.D. Major improvements in the cultivation of rice revolutionized Chinese agriculture. The control of water through hydraulic engineering (dams, ditches, dikes, etc.) allowed the drainage and irrigation of lands. The iron plow was introduced in the sixth century B.C. Chinese agriculture learned to use new fertilizers and chemically based pest control agents. The Chinese led the Europeans by 1500 years in the use of blast furnaces, allowing them to use cast iron and refine wrought iron from pig iron. Other notable (but by no means exhaustive) list of inventions include the invention of the compass, gunpowder, multistage rockets, paper, and the most technologically advanced ships of the time.

The extent of technological development in China by 1300 A.D. was apparently such that "China came within a hair's breadth of industrializing in the fourteenth century." (Jones, 1981, pg. 160). And yet, industrialization did not occur. In fact, by about 1400 A.D., technological progress in China slowed down dramatically to the point of almost ceasing altogether and in some cases even regressing. By the end of the Middle ages, Europe became the world's technological leader.

13.2.3 The Renaissance and Baroque Periods (1450 A.D. - 1750 A.D.)

Europe is unlike China in that it was never politically unified. Consequently, the centers of influence shifted over time across regions (or countries). The Renaissance (rebirth) era (1450 A.D. - 1600 A.D.) originated in the great Italian city states of Venice, Genoa and Florence with Spain (and to a lesser extent, Portugal) emerging as the world's leading power in the sixteenth century. With Spain's gradual decline following the defeat of the *Armada* (1588) at the hands of the English, the balance of power shifted in the Baroque era (1600 A.D. - 1750 A.D.) toward England, France and Holland.

This episode in European history is known more for art and exploration than for its technological achievements. Technological progress did occur throughout the period, but manifested itself more in the form of sequences of improvements to existing technology. Much of the growth in output were due to an expansion in international trade, facilitated by financial market innovations like commercial banking, as well as innovations in shipbuilding. In agriculture, the so-called new husbandry (the introduction of new crops, stall feeding of cattle, and the elimination of fallowing) led to major productivity gains. In energy use, medieval techniques were improved but not revolutionized, although the slow expansion of coal power was to have significant future ramifications. Major improvements were made in the efficiency of blast furnaces, mining techniques and hydraulic engineering. However, the diffusion of technology occurred very slowly, with old techniques coexisting along side best-practice techniques for centuries.

On the whole, the period was characterized more by its scientific achievements than its technological breakthroughs. The most famous visionary of the Renaissance, Leonardo da Vinci, designed thousands of machines and mechanical devices, including the helicopter, the retractable landing gear, the tank, and the parachute. Most of his ideas, however, defied any practical application until centuries later.

In the field of science, the Renaissance differed from Classical times in that the dichotomy between thinkers and makers broke down. Many scientists made their own instruments and contributed to the solution of practical problems associated with their manufacture. Galileo, for example, built his own telescopes and worked part time fixing machines. But the number of truly important technological breakthroughs brought about by great scientists still remained relatively small.

13.3 Thomas Malthus

Thomas Malthus (1766-1834) is probably most famous for his provocative work entitled *Essay on the Principle of Population* (1798), which was directed at

William Godwin (1756-1836) whose *Enquiry Concerning Political Justice* (1793) argued in favor of a more egalitarian society in order to end poverty. Godwin was a proponent of government welfare programs like the English Poor Law of 1601, which provided direct grants to the poor financed by property taxes. On the matter of the Poor Laws, Malthus had this to say:

> "The poor-laws of England tend to depress the general condition of the poor in these two ways. This first obvious tendency is to increase population without increasing the food for its support. A poor man may marry with little or no prospect of being able to support a family in independence. They may be said therefore in some measure to create the poor which they maintain; and as the provisions of the country must, in consequence of the increased population, be distributed to every man in smaller proportions, it is evident that the labour of those who are not supported by parish assistance will purchase a smaller quantity of provisions than before, and consequently, more of them must be driven to ask for support. Secondly, the quantity of provisions consumed in workhouses upon a part of the society, that cannot in general be considered the most valuable part, diminishes the shares that would otherwise belong to more worthy members; and thus in the same manner forces more to become dependent. ...
>
> Other circumstances being the same, it may be affirmed that countries are populous according to the quantity of human food which they produce or can acquire, and happy according to the liberality with which this food is divided, or the quantity which a day's labour will purchase. Corn countries are more populous than pasture countries, and rice countries more populous than corn countries. But their happiness does not depend upon their being thinly or fully inhabited, upon their poverty or their richness, their youth or their age, but on the proportion which the population and the good bear to each other."

With the publication of his *Essay*, controversial debates swept throughout England. To quote Malthus' biographer, James Bonar:

> "For thirty years it rained refutations. Malthus was the most abused man of the age, put down as a man who defended smallpox, slavery and child murder, who denounced soup kitchens, early marriage and parish allowances; who had the impudence to marry after preaching against the evils of a family; who thought the world so badly governed that the best actions do the most harm; who, in short, took all romance out of life."

The historical setting, in which Malthus brought out his work, must be considered. The poor, especially those in the rural areas, were numerous and were generally in a bad state. It was generally thought that the plight of the poor was due to the landed aristocracy, that they had the government levers in their hands and used them to advance the upper classes at the expense of the poor.

In contrast, Malthus explained the existence of the poor in terms of two 'unquenchable passions:' (1) the hunger for food; and (2) the hunger for sex. The only checks on population growth were wars, pestilence, famine, and 'moral restraints' (the willingness to refrain from sex). From these hungers and checks, Malthus reasoned that the population increases in a geometric ratio, while the means of subsistence increases in an arithmetic ratio. The most disturbing aspect of his theory was the conclusion that well-intentioned programs to help the poor would ultimately manifest themselves in the form of a greater population, leaving per capita incomes at their subsistence levels. It was this conclusion that ultimately led people to refer to economics as 'the dismal science.'

13.3.1 The Malthusian Growth Model

The Malthusian 'growth model' can be formalized in the following way. There are two key ingredients to his theory: (1) a technology for production of output; and (2) a technology for the production of people. The first technology can be expressed as an aggregate production function:

$$Y_t = F(K, N_t), \tag{13.1}$$

where Y_t denotes total real GDP, K denotes a fixed stock of capital (i.e., land), and N_t denotes population (i.e., the workforce of peasants). The production function F exhibits constant returns to scale in K and N_t. For example, suppose that F is a Cobb-Douglas function so that $F(K,N) = K^{1-\theta}N^{\theta}$, where $0 < \theta < 1$.

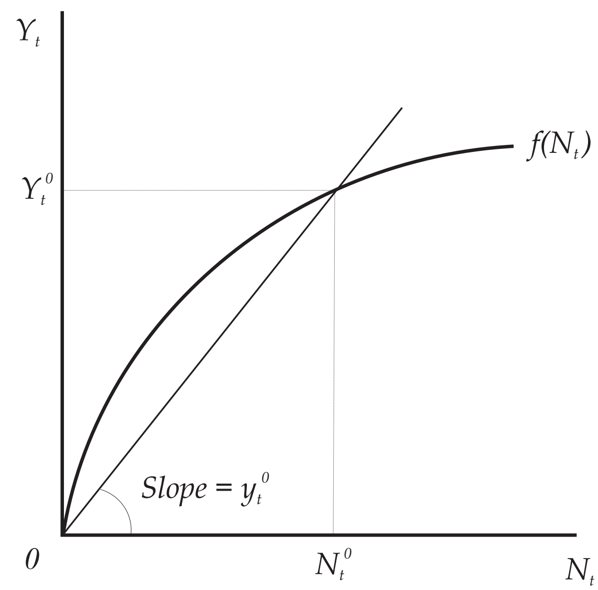
Because F exhibits constant returns to scale, it follows that per capita income $y_t \equiv Y_t/N_t$ is an increasing function of the capital-labor ratio. Since capital (land) is assumed to be in fixed supply, it follows that any increase in the population will lead to a lower capital-labor ratio, and hence a lower level of per capita output. Using our Cobb-Douglas function,

$$y_t = \frac{Y_t}{N_t} = \left(\frac{K}{N_t}\right)^{1-\theta},$$

which is clearly decreasing in N_t. The 'hunger for food' is captured by the assumption that all output is consumed.

On the other hand, the *total* GDP is clearly an increasing function of N_t; i.e., $Y_t = K^{1-\theta}N_t^{\theta}$. However, since land is fixed in supply, total output increases at a decreasing rate with the size of the population. Let $f(N_t) \equiv K^{1-\theta}N_t^{\theta}$ denote total output; this production function is depicted in Figure 13.1.

FIGURE 13.1
Malthusian Production Function



The technology for producing people is expressed as follows. First, assume that there is an exogenous birth rate $b > 0$. This assumption capture's Malthus' view that the rate of procreation is determined largely by noneconomic factors (such as the passion for sex). On the other hand, the mortality rate (especially among infants and the weaker members of society) was viewed by Malthus as determined in part by economic factors, primarily the level of material well-being as measured by per capita income y_t. An increase in y_t was thought to lower mortality rates (e.g., better fed babies are healthier and are less likely to die). Likewise, a decrease in y_t was thought to increase mortality rates. The dependence of the mortality rate m_t on living standards y_t can be expressed with the function:

$$m_t = m(y_t),$$

where $m(.)$ is a decreasing function of y_t.

Let n_t denote the (net) population growth rate; i.e., $n_t \equiv b - m_t$. Then it is clear that the population growth rate is an increasing function of living standards; a relation that we can write as:

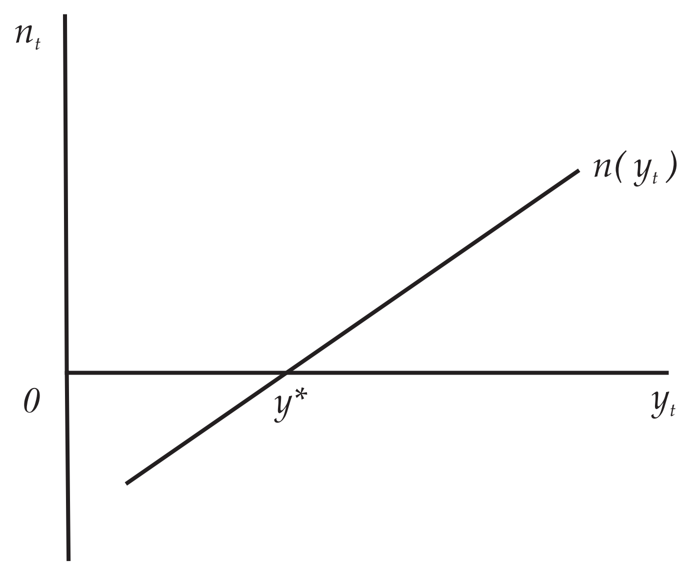
$$n_t = n(y_t), \tag{13.2}$$

where $n(.)$ is an increasing function of y_t. This relationship is depicted in Figure 9.2. It follows then that the total population N_t grows according to:

$$N_{t+1} = [1 + n(y_t)]N_t. \tag{13.3}$$

Note that when $y_t = y^*$ (in Figure 13.2), the net population growth rate is equal to zero (the birth rate is equal to the mortality rate) and the population stays constant.

FIGURE 13.2
Population Growth Rate



13.3.2 Dynamics

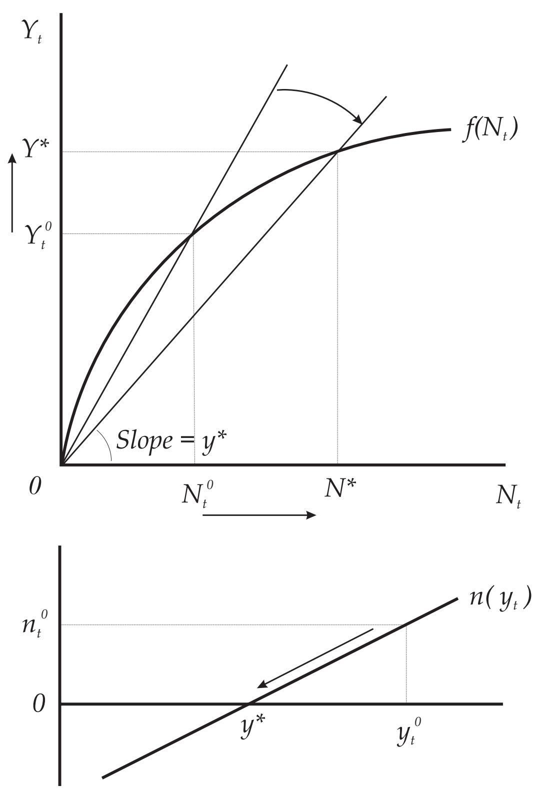
The Malthusian growth model has implications for the way real per capita GDP evolves over time, given some initial condition. The initial condition is given by the initial size of the population N_0. For example, suppose that N_0 is such that $y_0 = f(N_0) > y^*$, where y^* is the 'subsistence' level of income depicted in Figure 13.2. Thus, initially at least, per capita incomes are above subsistence levels.

According to Figure 13.2, if per capita income is above the subsistence level, then the population grows in size (the mortality rate is lower than the birth rate); i.e., $n_0 > 0$. Consequently, $N_1 > N_0$. However, according to Figure 13.1, the added population (working the same amount of land) leads to a reduction in living standards (the average product of labor falls); i.e., $y_1 < y_0$.

Since living standards are lower in period 1, Figure 13.2 tells us that mortality rates will be higher, leading to a decline in the population growth rate; i.e., $n_1 < n_0$. However, since the population growth rate is still positive, the population will continue to grow (although at a slower rate); i.e., $N_2 > N_1$. Again, referring to Figure 13.1, we see that the higher population continues to put pressure on the land, leading to a further decline in living standards; i.e., $y_2 < y_1$.

By applying this logic repeatedly, we see that per capita income will eventually (the process could take several years or even decades) converge to its subsistence level; i.e., $y_t \searrow y^*$. At the same time, total GDP and population will rise to higher 'long run' values; i.e., $Y_t \nearrow Y^*$ and $N_t \nearrow N^*$. These 'long run' values are sometimes referred to as 'steady states.' Figure 13.3 depicts these transition dynamics.

FIGURE 13.3
Transition Dynamics



- **Exercise 13.1.** Using a diagram similar to Figure 13.3, describe the dynamics that result when the initial population is such that $N_0 > N^*$.

13.3.3 Technological Progress in the Malthus Model

We know that Medieval Europe (800 - 1400 A.D.) did experience a considerable amount of technological progress and population growth (e.g., the population roughly doubled from 800 A.D. to 1300 A.D.). Less is known about how living standards changed, but there appears to be a general view that at least moderate improvements were realized.

We can model an exogenous technological advance (e.g., the invention of the wheelbarrow) as an outward shift of the aggregate production function. Let us assume that initially, the economy is in a steady state with living standards equal to y^*. In the period of the technology shock, per capita incomes rise as the improved technology makes the existing population more productive; i.e., $y_1 > y^*$. However, since living standards are now above subsistence levels, the population begins to grow; i.e., $N_1 > N_0$. Using the same argument described in the previous section, we can conclude that after the initial rise in per capita income, living standards will gradually decline back to their original level. In the meantime, the total population (and total GDP) expands to a new and higher steady state.

- **Exercise 13.2.** Using a diagram similar to Figure 13.3, describe the dynamics that result after the arrival of a new technology. Is the Malthus model consistent with the growth experience in Medieval Europe? Explain.

13.3.4 An Improvement in Health Conditions

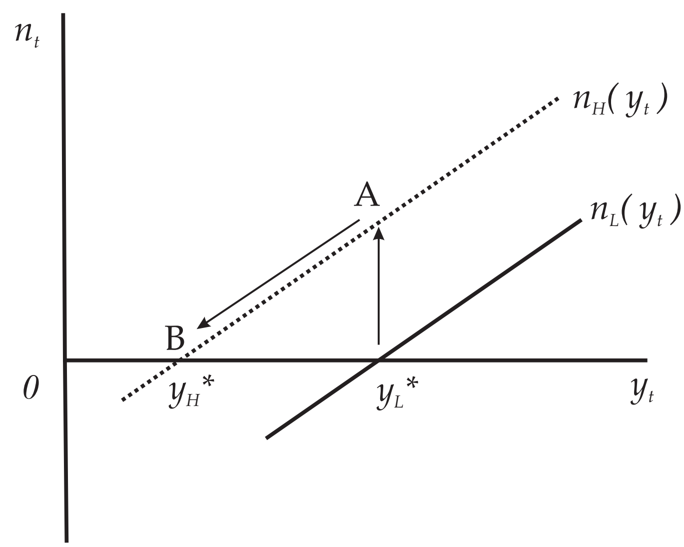
The number one cause of death in the history of mankind has not been war, but disease.[2] Slowly, medical science progressed to the point of identifying the primary causes of various diseases and recommending preventative measures (such as boiling water). For example, during the 1854 Cholera epidemic in London, John Snow (who had experienced the previous epidemics of 1832 and 1854) became convinced that Cholera was a water-borne disease (caused by all the human waste and pollution being dumped into the Thames river). Public works projects, like the Thames Embankment (which was motivated more by Parliamentarians' aversion to the 'Great Stink' emanating from the polluted Thames, than by concerns over Cholera), led to greatly improved health conditions and reduced mortality rates.

We can model a technological improvement in the 'health technology' as an upward shift in the function $n(y_t)$; i.e., a decline in the mortality rate associated with any given living standard y_t. Again, assume that the economy is initially at a steady state y^*, depicted as point A in Figure 13.4. The effect of such a change is to immediately reduce mortality rates which, according to Maltusian reasoning, then leads to an increase in population. But as the population expands, the effect is to reduce per capita income. Eventually, per capita incomes fall to a new and lower subsistence level y_N^*, depicted by point B in Figure 13.4.

[2] Even during wars, most soldiers evidently died from disease rather than combat wounds. For an interesting account of the role of disease in human history, I would recommend reading Jared Diamond's book *Guns, Germs and Steel* (1997).

That is, while the improved health conditions have the short run effect of lowering mortality rates, the subsequent decline in per capita reverses the effect so that in the long run, people are even worse off than before!

FIGURE 13.4
An Improvement in Health Technology



- **Exercise 13.3.** In 1347, the population of Europe was around 75 million. In that year, the continent was ravaged by a bubonic plague (the Black Death), which killed approximately 25 million people over a five year period (roughly one-third of the population). The ensuing labor shortages apparently led to a significant increase in real wages (per capita incomes), although total output fell. Using a diagram similar to Figure 13.4, describe the dynamics for per capita income in the Malthusian model when the economy is subject to a transitory increase in the mortality rate.

13.3.5 Confronting the Evidence

For most economies prior to 1800, growth in real per capita incomes were moderate to nonexistent. Since 1800, most economies have exhibited at least some growth in per capita incomes, but for many economies (that today comprise the world's underdeveloped nations), growth rates have been relatively low, leaving their per capita income levels far behind the leading economies of the world.

The Maltusian model has a difficult time accounting for the sustained increase in per capita income experienced by many countries since 1800, especially in light of the sharp declines in mortality rates that have been brought about by continuing advancements in medical science. It is conceivable that persistent declines in the birth rate offset the declines in mortality rates (downward shifts of the population growth function in Figure 13.2) together with the continual appearance of technological advancements together could result in long periods of growth in per capita incomes. But the birth rate has a lower bound of zero and in any case, while birth rates do seem to decline with per capita income, most advanced economies continue to exhibit positive population growth.

In accounting for cross-section differences in per capita incomes, the Malthusian model suggests that countries with high population densities (owing to high birth rates) will be those economies exhibiting the lowest per capita incomes. One can certainly find modern day countries, like Bangladesh, that fit this description. On the other hand, many densely populated economies, such as Hong Kong, Japan and the Netherlands have higher than average living standards. As well, there are many cases in which low living standards are found in economies with low population density. China, for example, has more than twice as much cultivated land per capita as Great Britain or Germany.

At best, the Malthusian model can be regarded as giving a reasonable account of the pattern of economic development in the world prior to the Industrial Revolution. Certainly, it seems to be true that the vast bulk of technological improvements prior to 1800 manifested themselves primarily in the form of larger populations (and total output), with only modest improvements in per capita incomes.

13.4 Fertility Choice

The Malthusian model does not actually model the fertility choices that households make. The model simply assumes that the husband and wife decide to create children for really no reason at all. Perhaps children are simply the by-product of uncontrollable passion or some primeval urge to propagate one's genetic material. Or perhaps in some cultures, men perceive that their status is enhanced with prolific displays of fertility. Implicitly, it is assumed that the fertility choices that people make are 'irrational.' In particular, some simple family planning (choosing to reduce the birth rate) would appear to go a long way to improving the living standards of future generations.

- **Exercise 13.4.** Suppose that individuals could be taught to choose the birth rate according to: $b(y) = m(y)$ (i.e., to produce just enough children to replace those people who die). Explain how technological progress would now result in higher per capita incomes.

While it is certainly the case that the family planning practices of some households seem to defy rational explanation, perhaps it is going too far to suggest that the majority of fertility choices are made largely independent of economic considerations. In fact, it seems more likely to suppose that fertility is a rational choice, even in lesser developed economies. A 1984 World Bank report puts it this way (quoted from Razin and Sadka, 1995, pg. 5):

> All parents everywhere get pleasure from children. But children involve economic costs; parents have to spend time and money bringing them up. Children are also a form of investment–providing short-term benefits if they work during childhood, long-term benefits if they support parents in old age. There are several good reasons why, for poor parents, the economic costs of children are low, the economic (and other) benefits of children are high, and having many children makes economic sense.

Here, I would like to focus on the idea of children as constituting a form of investment. What appears to be true of many primitive societies is a distinct lack in the ability for large segments of society to accumulate wealth in the form of capital goods or (claims to) land. Partly this was due to a lack of well-developed financial markets and partly this was due to the problem of theft (only the very rich could afford to spend the resources necessary to protect their property). Given such constraints, in may well make sense for poorer families to store their wealth through other means, for example, by investing in children (although, children can also be stolen, for example, by conscription into the military or by the grim reaper).

Let us try to formalize this idea by way of a simple model. Consider an economy in which time evolves according to $t = 0, 1, 2, ..., \infty$. For simplicity, assume that individuals live for two periods. In period one they are 'young' and in period two they are 'old.' Let $c_t(j)$ denote the consumption enjoyed by an individual at period t in the j^{th} period of life, where $j = 1, 2$. Assume that individuals have preferences defined over their lifetime consumption profile $(c_t(1), c_{t+1}(2))$, with:

$$U_t = \ln c_t(1) + \beta \ln c_{t+1}(2),$$

where $0 < \beta < 1$ is a subjective discount factor. For these preferences, the marginal rate of substitution between time-dated consumption is given by $MRS = c_{t+1}(2)/(\beta c_t(1))$.

Let N_t denote the number of young people alive at date t, so that N_{t-1} represents the number of old people alive at date t. The population of young people grows according to:

$$N_{t+1} = n_t N_t,$$

where n_t here is the *gross* population growth rate; i.e., the average number of children per (young) family. Note that $n_t > 1$ means that the population is expanding, while $n_t < 1$ means that the population is contracting. We will assume that n_t is chosen by the young according to some rational economic principle.

Assume that only the young can work and that they supply one unit of labor at the market wage rate w_t. Because the old cannot work and because the have no financial wealth to draw on, they must rely on the current generation of young people (their children) to support them. Suppose that these intergenerational transfers take the following simple form: The young set aside some fraction $0 < \theta < 1$ of their current income for the old. Since the old at date t have n_{t-1} children, the old end up consuming:

$$c_t(2) = n_{t-1}\theta w_t. \tag{13.4}$$

This expression tells us that the living standards of old people are an increasing function of the number of children they have supporting them. As well, their living standards are an increasing function of the real wage earned by their children.

Creating and raising children entails costs. Assume that the cost of n_t children is n_t units of output. In this case, the consumption accruing to a young person (or family) at date t is given by:

$$c_t(1) = (1 - \theta)w_t - n_t. \tag{13.5}$$

By substituting equation (13.5) into equation (13.4), with the latter equation updated one period, we can derive the following lifetime budget constraint for a representative young person:

$$c_t(1) + \frac{c_{t+1}(2)}{\theta w_{t+1}} = (1 - \theta)w_t. \tag{13.6}$$

Equation (13.6) should look familiar to you. In particular, the left hand side of the constraint represents the present value of lifetime consumption spending. But instead of discounting future consumption by the interest rate (which does not exist here since there are no financial markets), future consumption is discounted by a number that is proportional to the *future* wage rate. In a sense, the future wage rate represents the implicit interest rate that is earned from investing in children today. Figure 13.5 displays the optimal choice for a given pattern of wages (w_t, w_{t+1}).

FIGURE 13.5
Optimal Family Size

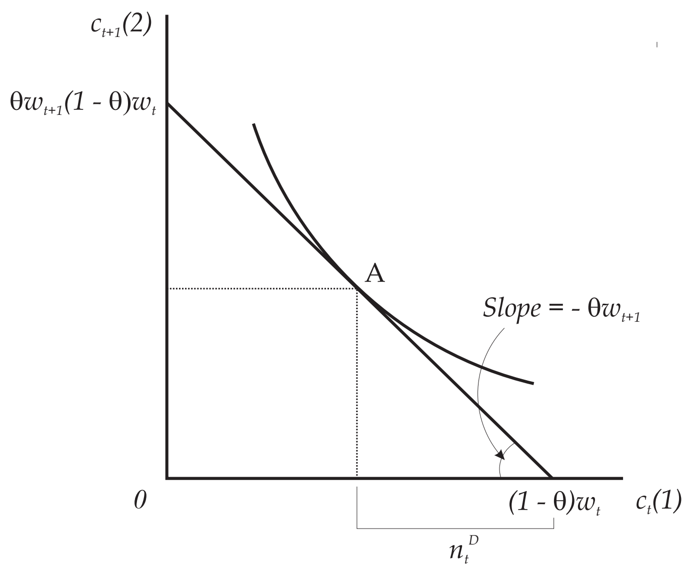


Figure 13.5 makes clear the analog between the savings decision analyzed in Chapter 4 and the investment choice in children as a vehicle for saving in the absence of any financial market. While having more children reduces the living standards when young, it increases living standards when old. At point A, the marginal cost and benefit of children are exactly equal. Note that the desired family size generally depends on both current and future wages; i.e., $n_t^D = n^D(w_t, w_{t+1})$.

- **Exercise 13.5.** How does desired family size depend on current and future wages? Explain.

We will now explain how wages are determined. Assume that the aggregate production technology is given by (13.1). The fixed factor K, which we interpret to be land, is owned by a separate class of individuals (landlords). Imagine that landlords are relatively few in number and that they form an exclusive club (so that most people are excluded from owning land). Landowners hire workers at the competitive wage rate w_t in order to maximize the return on their land $D_t = F(K, N_t) - w_t N_t$. As in Appendix 2.A, the profit maximizing labor input $N_t^D = N^D(w_t)$ is the one that just equates the marginal benefit of labor (the marginal product of labor) to the marginal cost (the wage rate); i.e.,

$$MPL(N^D) = w_t.$$

The equilibrium wage rate w_t^* is determined by equating the supply and demand for labor; i.e.,

$$N^D(w_t^*) = N_t.$$

Alternatively, you should be able to show that the equilibrium wage rate can also be expressed as: $w_t^* = MPL(N_t)$.

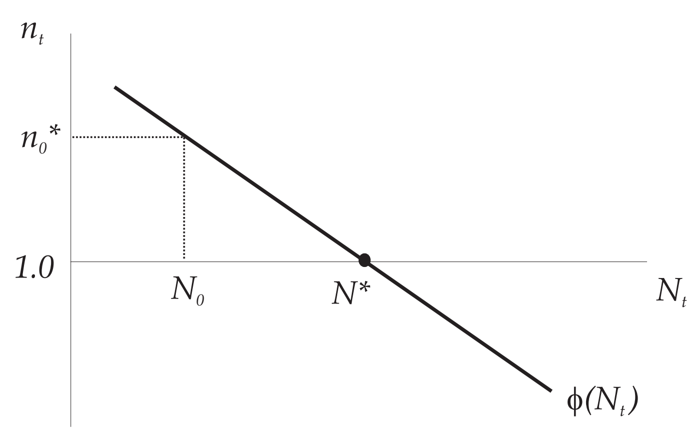
- **Exercise 13.6.** How does the equilibrium wage rate depend on the supply of labor N_t? Explain.

Mathematically, the general equilibrium of our model economy is characterized by the following condition:

$$N_{t+1} = n^D(w_t^*, w_{t+1}^*)N_t$$

where $w_t^* = MPL(N_t)$, with N_t given as of period t. These expressions implicitly defines a function $n_t^* = \phi(N_t)$.[3] A stable steady state will exist if $\phi'(N) < 0$, so let us make this assumption here.[4] This condition asserts that the equilibrium population growth rate is a decreasing function of population size; see Figure 13.6.

FIGURE 13.6
Equilibrium Population Dynamics



[3]The function ϕ is defined implicitly by:

$$\phi(N_t) = n^D(MPL(N_t), MPL(\phi(N_t)N_t).$$

[4]I am pretty sure that for sufficiently large populations, the function ϕ must eventually decline with population since land is in fixed supply.

In Figure 13.6, the initial population of young people is given by N_0, which results in population growth. In the subsequent period, $N_1 > N_0$, which puts added pressure on the limited supply of land (just as in the Malthusian model), resulting in a decline in the equilibrium wage.[5] Unlike the Malthusian model, however, people here are making rational choices about family size. As the population expands, it is rational to reduce family size. Eventually, the population reaches a steady state level N^*.

From the condition that characterizes optimal family size (Figure 9.5), the steady state consumption levels must satisfy:

$$\frac{1}{\beta}\frac{c^*(1)}{c^*(2)} = \theta w^* = \theta MPL(N^*).$$

Now, assume that the share parameter θ is chosen according to a principle of 'long-run fairness,' so that $c^*(1) = c^*(2)$. Notice that this does not necessarily imply that consumption is equated across generations during the transition to a steady state; it only implies that consumption is equated in the steady state. In this case, the equilibrium steady state population size (and wage rate) is determined by:

$$MPL(N^*) = \frac{1}{\theta\beta} = w^*. \tag{13.7}$$

Condition (13.7) tells us that the steady state population N^* is determined by the nature of the production technology (MPL) and the nature of preferences (β).

Now, beginning in some steady state, let us examine how this economy reacts to a technology shock that improves production methods. The effect of the shock is to increase the marginal product of labor at every level of employment (so that the function ϕ in Figure 9.6 shifts upward). From equation (13.7), we see that the initial effect of the shock is to increase the wage rate above w^*. A standard consumption-smoothing argument suggests that consumption rises initially for both the young and old. The way that the initial young can guarantee higher future consumption is by having more children (re: Exercise 9.5). The increasing population, however, puts downward pressure on the wage until it eventually falls to its initial value w^*. From equation (13.7), we see that the long-run wage rate depends only on β and not on the nature of technology. It follows, therefore, that the long-run living standards of those individuals who must save by investing in children remains unaffected by technological progress.

The effect of technological progress on per capita income depends on the breeding habits of landlord families and the relative importance of land versus labor in the production process. If landlord family size remains constant over time, then per capita income will rise since the shock increases the return to land. But if land accounts for a relatively small fraction of total output, then the impact on per capita GDP will be small.

[5] As the future wage rate is expected to decline, the return to investing in children also declines (the substitution effect) would further curtail the production of children.

13.4.1 Policy Implications

The basic point of this analysis is to show that what appears (to us) to be irrational family planning may in fact be the consequence of rational choices made by individuals who are prevented from saving through the accumulation of financial assets or physical capital. The policy implications here differ quite radically from those that one might deduce from the Malthusian model. In particular, the Malthusian model suggests that a government program designed to limit the breeding rate of peasants might be a good idea. An example of this is the 'one-child' policy implemented by the Chinese government in 1980.[6] In contrast, the model developed in this section suggests that a better idea might be to make participation in capital markets more accessible for the poor. Less reliance on children to finance retirement living standards would imply lower population growth rates and higher material living standards for all people.

[6] http:// nhs.needham.k12.ma.us/ cur/ kane98/ kanep2/ chinas1kid/ dcva2.html

13.5 Problems

1. Many countries have implemented *pay-as-you-go* (PAYG) public pension systems. A PAYG system taxes current income earnings (the young) and transfers these resources to the initial old. Explain how such a system could also serve to reduce population growth.

2. Many countries with PAYG pension systems are currently struggling with the problem of population growth rates that are *too low*; e.g., see: www.oecdobserver.org/ news/ fullstory.php/ aid/563/ Can_ governments_ influence_ population_ growth_ .html, for the case of Sweden. Use the model developed in this section to interpret this phenomenon.

3. Economists have advocated replacing the PAYG pension system with a *fully funded* (FF) pension system. Whereas the PAYG system transfers resources across generations (from the young to the old in perpetuity), the FF system taxes the young and invests the proceeds in capital markets (so that there are no intergenerational transfers). Does the FF system sound like a good idea? Explain.

13.6 References

1. Diamond, Jared (1997). *Guns, Germs and Steel: The Fates of Human Societies*, New York: W.W. Norton.

2. Godwin, William (1793). *Enquiry Concerning Political Justice*, http:// web.bilkent.edu.tr/ Online/www.english.upenn.edu /jlynch/ Frank/ Godwin/ pjtp.html

3. Jones, Eric L. (1981). *The European Miracle*, Cambridge: Cambridge University Press.

4. Malthus, Thomas (1798). *Essay on the Principle of Population*, www.ac.wwu.edu/ ~stephan/ malthus/ malthus.0.html

5. Mokyr, Joel (1990). *The Levers of Riches: Technological Creativity and Economic Progress*, Oxford University Press, New York.

6. Razin, Assaf and Efraim Sadka (1995). *Population Economics*, The MIT Press, Cambridge, Massachussetts.

Chapter 14

Modern Economic Development

14.1 Introduction

In the previous chapter, we saw that despite the fact of technological progress throughout the ages, material livings standards for the average person changed relatively little. It also appears to be true that differences in material living standards across countries (at any point in time) were relatively modest. For example, Bairoch (1993) and Pomeranz (1998) argue that living standards across countries in Europe, China, and the Indian subcontinent were roughly comparable in 1800. Parente and Prescott (1999) show that material living standards in 1820 across the 'western' world and 'eastern' world differed only by a factor of about 2. Overall, the Malthusian growth model appears to account reasonably well for the pattern of economic development for much of human history.

But things started to change sometime in the early part of the 19th century, around the time of the Industrial Revolution that was occurring (primarily in Great Britain, continental Europe, and later in the United States). There is no question that the pace of technological progress accelerated during this period. The list of technological innovations at this time are legendary and include: Watt's steam engine, Poncelet's waterwheel, Cort's puddling and rolling process (for iron manufacture), Hargeave's spinning jenny, Crompton's mule, Whitney's cotton gin, Wilkensen's high-precision drills, Lebon's gas light, Montgolfiers' hydrogen balloon, and so on. The technological innovations in the British manufacturing sector increased output dramatically. For example, the price of cotton declined by 85% between 1780 and 1850. At the same time, per capita incomes in the industrialized countries began to rise measurably for the first time in history.

It is too easy (and probably wrong) to argue that the innovations associated with the Industrial Revolution was the 'cause' of the rise in per capita income in the western world. In particular, we have already seen in Chapter 13 that technological progress does not in itself guarantee rising living standards. Why, for example, did the rapid pace of technological development simply not dissipate itself entirely in the form of greater populations, consistent with historical patterns?[1] Clearly, something else other than just technological progress must be a part of any satisfactory explanation.

As per capita incomes began to grow rapidly in countries that became industrialized (i.e., primarily the western world), living standards in most other countries increased at a much more modest pace. For the first time in history, there emerged a large and growing disparity in the living standards of people across the world. For example, Parente and Prescott (1999) report that by 1950, the disparity in real per capita income across the 'west' and the 'east' grew to a factor of 7.5; i.e., see Table 14.1.

Table 14.1			
Per Capita Income (1990 US$)			
Year	West	East	West/East
1820	1,140	540	2.1
1870	1,880	560	3.3
1900	2,870	580	4.2
1913	3,590	740	4.8
1950	5,450	727	7.5
1973	10,930	1,670	6.5
1989	13,980	2,970	4.7
1992	13,790	3,240	4.3

The data in Table 10.1 presents us with a bit of a puzzle: why did growth in the east (as well as many other places on the planet) lag behind the west for so many decades? Obviously, most of these countries did not industrialize themselves as in the west, but the question is why not? It seems hard to believe that people living in the east were unaware of new technological developments or unaccustomed to technological progress. After all, as was pointed out in the previous chapter, most of the world's technological leaders have historically been located in what we now call the east (the Moslem world, the Indian subcontinent, China). At the same time, it is interesting to note that the populations in the eastern world exploded over this time period (in accord with the Malthusian model).

Some social scientists (notably, those with a Marxian bent) have laid the blame squarely on the alleged exploitation undertaken by many colonial powers (e.g., Great Britain in Africa). But conquest and 'exploitation' have been with us throughout human history and has a fine tradition among many eastern

[1] While populations did rise in the west, total income rose even faster.

cultures too. So, perhaps one might ask why the east did not emerge as the world's colonial power?

In any case, it simply is not true that all eastern countries were under colonial domination. For example, Hong Kong remained a British colony up until 1997 while mainland China was never effectively controlled by Britain for any length of time. And yet, while Hong Kong and mainland China share many cultural similarities, per capita incomes in Hong Kong have been much higher than on the mainland over the period of British 'exploitation.' Similarly, Japan was never directly under foreign influence until the end of the second world war. Of course, this period of foreign influence in Japan happens to coincide with a period of remarkable growth for the Japanese economy.

Table 14.1 reveals another interesting fact. Contrary to what many people might believe, the disparity in per capita incomes across many regions of the world appear to be diminishing. A large part of this phenomenon is attributable to the very rapid growth experienced recently by economies like China, India and the so-called 'Asian tigers' (Japan, South Korea, Singapore, Taiwan). So again, the puzzle is why did (or have) only some countries managed to embark on a process of 'catch up' while others have been left behind? For example, the disparity in incomes across the United States and some countries in the sub-Saharan African continent are still different by a factor of 30!

The 'development puzzle' that concerns us can be looked at also in terms of countries within the so-called western world. It is not true, for example, that all western countries have developed at the same pace; see, for example, Figure 14.1. The same can be said of different regions within a country. For example, why are eastern Canadian provinces so much poorer than those in central and western Canada? Why is the south of Italy so much poorer than the north? Why is the northern Korean peninsula so much poorer than the South (although, these are presently separate countries)? In short, what accounts for the vast disparity in per capita incomes that have emerged since the Industrial Revolution?

FIGURE 14.1

Real per Capita GDP Relative to the United States

Selected Countries

14.2 The Solow Model

The persistent rise in living standards witnessed in many countries and the large disparity in per capita incomes across countries are facts that are difficult to account for with the Malthusian model. For this reason, economists turned to developing an alternative theory; one that would hopefully be more consistent with recent observation. The main model that emerged in the mid-20th century was the so-called Solow growth model, in honor of Robert Solow who formalized the basic idea in 1956 (Solow, 1956). Keep in mind that, like the Malthusian model, the Solow model does not actually explain why technological progress occurs; i.e., it treats the level (and growth rate) of technology as an exogenous variable. The name of the game here is see whether differences in per capita incomes can be explained by (exogenous) technological developments (among possibly other factors).

We remarked in Chapter 13 that China has much more cultivated land per capita as Great Britain. But what then accounts for the higher standard of living in Great Britain? One explanation is to be found in the fact that Great Britain has considerably more railroads, refineries, factories and machines per capita. In other words, Great Britain has more *physical capital* per capita (a higher capital-labor ratio) relative to China (and indeed, relative to Great Britain 100 years ago). Of course, this does not answer the question of *why* Britain has more physical capital than China. The model of endogenous fertility choices in Chapter 9 suggests that one reason might be that the institutional environment in China is (or was) such that the Chinese (like many lesser developed economies in history) are restricted from accumulating physical or financial assets, so that 'retirement saving' must be conducted through family size.[2]

In any case, the Solow model is firmly rooted in the model developed in Chapter 6, which assumes that individuals can save through the accumulation of physical and/or financial assets. Now, unlike land, which is largely in fixed supply (this is not exactly true, since new land can be cultivated), the supply of physical capital can grow with virtually no limit by producing new capital goods. Hence, the first modification introduced by the Solow model is the idea that output is produced with both labor and a time-varying stock of physical capital; i.e.,

$$Y_t = F(K_t, N_t), \tag{14.1}$$

and that the capital stock can grow with net additions of new capital. Let X_t denote gross additions to the capital stock (i.e., gross investment). Assuming that the capital stock depreciates at a constant rate $0 \leq \delta \leq 1$, the net addition to the capital stock is given by $X_t - \delta K_t$, so that the capital stock evolves according to:

$$K_{t+1} = K_t + X_t - \delta K_t. \tag{14.2}$$

[2] Of course, this does not explain why the institutional environments should differ the way that they do.

Of course, by allocating resources in an economy toward the construction of new capital goods (investment), an economy is necessarily diverting resources away from the production (and hence consumption) of consumer goods and services. In other words, individuals as a group must be *saving*. In the Malthus model, individuals were modeled as being either unwilling or unable to save. Perhaps this was a good description for economies prior to 1800, but is not a good description of aggregate behavior since then. Thus, the second modification introduced by the Solow model is the idea that a part of current GDP is saved; i.e.,

$$S_t = \sigma Y_t, \tag{14.3}$$

where $0 < \sigma < 1$ is the *saving rate*. In the Solow model, the saving rate is viewed as an exogenous parameter. However, as we learned in Chapter 6, the saving rate is likely determined by 'deeper' parameters describing preferences (e.g., time preference) and technology.

The final modification made by the Solow model is in how it describes population growth. Unlike the Malthus model, which assumed that mortality rates were a decreasing function of living standards, the Solow model simply assumes that the population growth rate is determined exogenously (i.e., is insensitive to living standards and determined largely by cultural factors). Consequently, the population grows according to:

$$N_{t+1} = (1+n)N_t, \tag{14.4}$$

where n denotes the *net* population growth rate.

We are now in a position to examine the implications of the Solow model. We can start with the production function in (14.1), letting $F(K,N) = K^{1-\theta}N^{\theta}$ as we did earlier. As before, we can define per capita output $y_t \equiv Y_t/N_t$ so that:

$$y_t = \left(\frac{K_t}{N_t}\right)^{1-\theta} \equiv f(k_t),$$

where $k_t \equiv K_t/N_t$ is the capital-labor ratio. Thus, per capita output is an increasing and concave function of the capital-labor ratio (try drawing this).

By dividing the saving function (14.3) through by N_t, we can rewrite it in per capita terms:

$$\begin{aligned} s_t &= \sigma y_t; \\ &= \sigma f(k_t). \end{aligned} \tag{14.5}$$

Now, take equation (14.2) and rewrite it in the following way:

$$\frac{N_{t+1}}{N_{t+1}}\frac{K_{t+1}}{N_t} = \frac{K_t}{N_t} + \frac{X_t}{N_t} - \frac{\delta K_t}{N_t}.$$

Using equation (14.4), we can then express this relation as:

$$(1+n)k_{t+1} = (1-\delta)k_t + x_t, \tag{14.6}$$

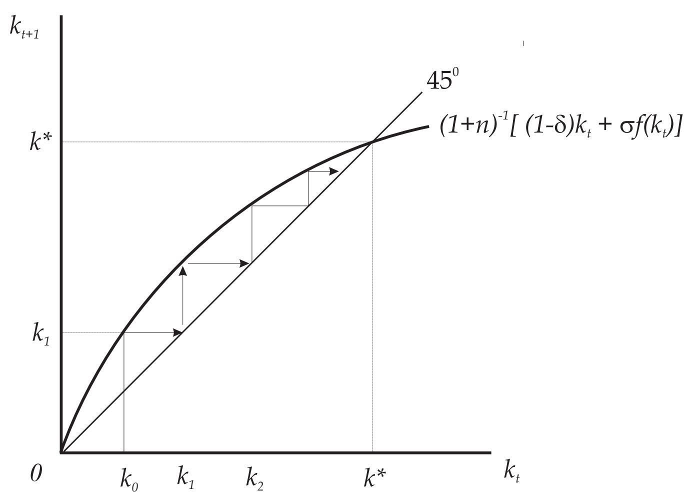
where $x_t \equiv X_t/N_t$.

In a closed economy, net saving must equal net investment; i.e., $s_t = x_t$. We can therefore combine equations (14.5) and (14.6) to derive:

$$(1+n)k_{t+1} = (1-\delta)k_t + \sigma f(k_t). \tag{14.7}$$

For any initial condition k_0, equation (14.7) completely describes the dynamics of the Solow growth model. In particular, given some k_0, we can use equation (14.7) to calculate $k_1 = (1+n)^{-1}[(1-\delta)k_0 + f(k_0)]$. Then, knowing k_1, we can calculate $k_2 = (+n)^{-1}[(1-\delta)k_1 + f(k_1)]$, and so on. Once we know how the capital-labor ratio evolves over time, it is a simple matter to calculate the time-path for other variables since they are all functions of the capital-labor ratio; e.g., $y_t = f(k_t)$.Equation (14.7) is depicted graphically in Figure 14.2.

FIGURE 14.2
Dynamics in the Solow Model



As in the Malthus model, we see from Figure 14.2 that the Solow model predicts that an economy will converge to a steady state; i.e., where $k_{t+1} = k_t = k^*$. The steady state capital-labor ratio k^* implies a steady state per capita income level $y^* = f(k^*)$. If the initial capital stock is $k_0 < k^*$, then $k_t \nearrow k^*$ and $y_t \nearrow y^*$. Thus, in contrast to the Malthus model, the Solow model predicts that real per capita GDP will *grow* during the transition period toward steady state, even as the population continues to grow. In the steady state, however, growth in per capita income ceases. Total income, however, will continue to grow at the population growth rate; i.e., $Y_t^* = f(k^*)N_t$.

Unlike the Malthus model, the Solow model predicts that growth in per capita income will occur, at least in the 'short run' (possibly, several decades) as the economy makes a transition to a steady state. This growth comes about because individuals save output to a degree that more than compensates for the depreciated capital and expanding population. However, as the capital-labor ratio rises over time, diminishing returns begin to set in (i.e., output per capita does not increase linearly with the capital-labor ratio). Eventually, the returns to capital accumulation fall to the point where just enough investment occurs to keep the capital-labor ratio constant over time.

The transition dynamics predicted by the Solow model may go some way to partially explaining the rapid growth trajectories experienced over the last few decades in some economies, for example, the 'Asian tigers' of southeast Asia (see Figure 14.1). Taken at face value, the explanation is that the primary difference between the U.S. and these economies in 1950 was their respective 'initial' capital stocks. While there may certainly be an element of truth to this, the theory is unsatisfactory for a number of reasons. For example, based on the similarity in per capita incomes across countries in the world circa 1800, one might reasonably infer that 'initial' capital stocks were not very different in 1800. And yet, some economies industrialized, while others did not. Transition dynamics may explain a part of the growth trajectory for those countries who chose to industrialize at later dates, but it does not explain the long delay in industrialization.

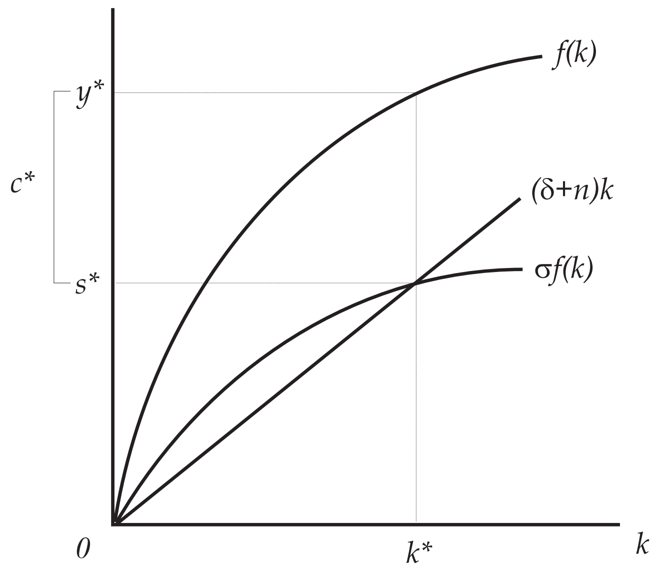
14.2.1 Steady State in the Solow Model

Because the level of income disparity has persisted for so long across many economies, it may make more sense to examine the steady state of the Solow model and see how the model interprets the source of 'long-run' differences in per capita income. By setting $k_{t+1} = k_t = k^*$, we see from equation (14.7) that the steady state capital-labor ratio satisfies:

$$\sigma f(k^*) = (n + \delta)k^*. \tag{14.8}$$

Equation (14.8) describes the determination of k^* as a function of σ, n, δ and f. Figure 10.3 depicts equation (14.8) diagrammatically.

FIGURE 14.3
Steady State in the Solow Model

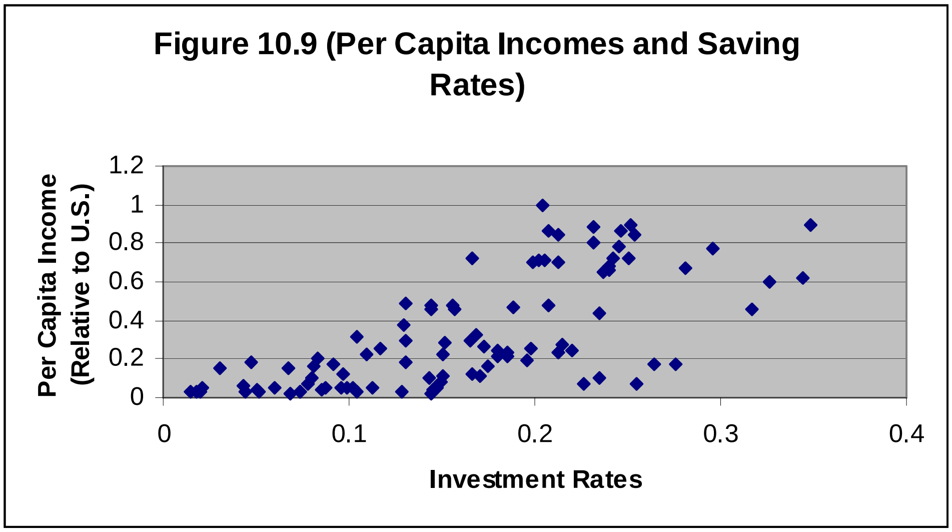


According to the Solow model, exogenous differences in saving rates (σ), population growth rates (n), or in technology (f) may account for differences in long-run living standards (y^*). Of course, the theory does not explain why there should be any differences in these parameters, but let's leave this issue aside for the moment.

14.2.2 Differences in Saving Rates

Using either Figure 14.2 or 14.3, we see that the Solow model predicts that countries with higher saving rates will have higher capital-labor ratios and hence, higher per capita income levels. The intuition for this is straightforward: higher rates of saving imply higher levels of wealth and therefore, higher levels of income.

Using a cross section of 109 countries, Figure 10.9 (mislabelled) plots the per capita income of various countries (relative to the U.S.) across saving rates (using the investment rate as a proxy for the saving rate). As the figure reveals, there appears to be a positive correlation between per capita income and the saving rate; a prediction that is consistent with the Solow model.



- **Exercise 14.1.** Consider an economy that is initially in a steady state. Imagine now that there is an exogenous increase in the saving rate. Trace out the dynamics for income and consumption predicted by the Solow model. If the transition period to the new steady state was to take several decades, would all generations necessarily benefit from the higher long-run income levels?

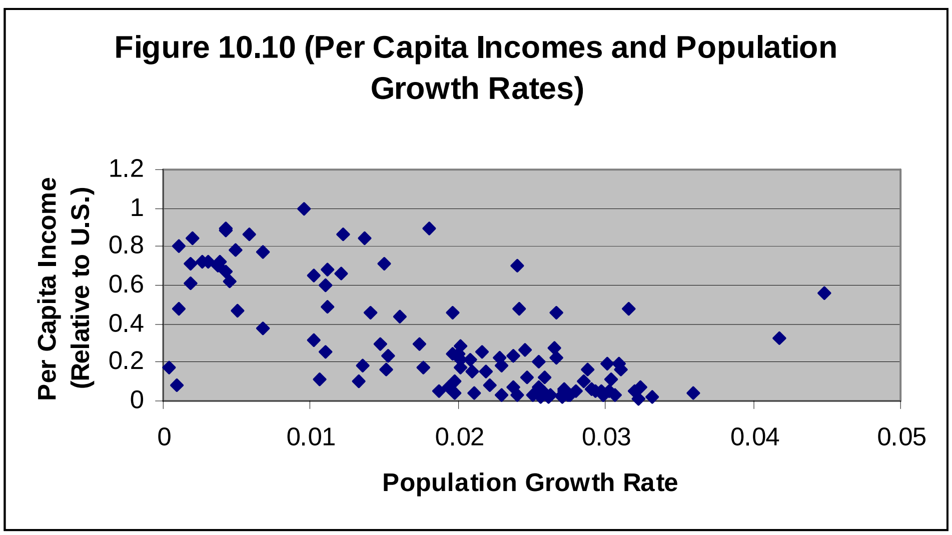
- **Exercise 14.2.** According to the Solow model, can sustained economic growth result from ever rising saving rates? Explain.

The deeper question here, of course, is why countries may differ in their rate of saving. One explanation may be that savings rates (or the discount factor in preferences) are 'culturally' determined. Explanations that are based on 'cultural' differences, however, suffer from a number of defects. For example, when individuals from many cultures arrive as immigrants to a new country, they often adopt economic behavior that is more in line with their new countrymen. As well, many culturally similar countries (like North and South Korea) likely have very different saving rates (I have not checked this out). A more likely explanation lies in the structure of incentives across economies. Some of these incentives are determined politically, for example, the rate at which the return to saving and investment is taxed. But then, the question simply turns to what determines these differences in incentives.

14.2.3 Differences in Population Growth Rates

Again, using either Figure 14.2 or 14.3, we see that the Solow model predicts that countries with relatively low population growth rates should enjoy relatively high per capita incomes. This prediction is similar to the Malthusian prediction, but different in an important way. In particular, the Malthusian model predicts that countries with high population densities should have lower living standards, while the Solow model makes no such prediction. That is, in the Solow model, even high density populations can enjoy high income levels, as long as they have a sufficient amount of physical capital. A higher population *growth rate*, however, spreads any given level of capital investment more thinly across a growing population, which is what accounts for their lower capital-labor ratios.

Using a cross section of 109 countries, Figure 10.10 (mislabelled) plots the per capita income of various countries (relative to the U.S.) across population growth rates. According to this figure, there appears to be a mildly negative correlation between per capita income and the population growth rate; a prediction that is also consistent with the Solow model.



Again, the deeper question here is why do different countries have different population growth rates? According to the Solow model, the population growth rate is exogenous. In reality, however, people make choices about how many children to have or where to move (i.e., population growth is endogenous). A satisfactory explanation that is based on differences in population growth rates should account for this endogeneity.

14.2.4 Differences in Technology

Parente and Prescott (1999) argue that exogenous differences in savings rates (and presumably population growth rates) can account for only a relatively small amount of the income disparity observed in the data. According to these authors, the proximate cause of income disparity is attributable to differences in total factor productivity (basically, the efficiency of physical capital and labor). In the context of the Solow model, we can capture differences in total factor productivity by assuming that different economies are endowed with different technologies f.

For example, consider two economies A and B that have technologies $f^A(k)$ and $f^B(k)$, where $f^A > f^B$ for all values of k. You can capture this difference by way of a diagram similar to Figure 14.3. Since the more productive economy can produce more output for any given amount of capital (per worker), national savings and investment will be higher. This higher level of investment translates into a higher steady state capital labor ratio. Per capita output will be higher in economy A for two reasons: First, it will have more capital (relative to workers); and second, it is more productive for any given capital-labor ratio.

Differences in the production technology f across countries can be thought of as differences in the type of engineering knowledge and work practices implemented by individuals living in different economies. Differences in f might also arise because of differences in the skill level or education of workers across countries. But while these types of differences might plausibly lead to the observed discrepancy in material living standards around the world, the key question remains unanswered: Why don't lesser developed economies simply adopt the technologies and work practices that have been so successful in the so-called developed world? Likewise, this theory does not explain why countries that were largely very similar 200 years ago are so different today (in terms of living standards).

14.3 The Politics of Economic Development

Perhaps more headway can be made in understanding the process of economic development by recognizing the politics that are involved in implementing new technologies and work practices. By 'politics' I mean the conflict that appears to arise among 'different' groups of people and the various methods by which people in conflict with one another use to try to protect themselves and/or harm others for private gain.

One thing that we should keep in mind about technological improvement is that it rarely, if ever, benefits everyone in an economy in precisely the same way. Indeed, some technological improvements may actually end up hurting some segments of the economy, even if the economy as a whole benefits in some sense. Whether a particular technological improvement can be implemented or not likely depends to some extent on the relative political power of those special interests that stand to benefit or lose from the introduction of a new technology. If the political power of the potential losers is strong enough, they may be able to erect laws (or vote for political parties that act on their behalf) that prevent the adoption of a new technology.

14.3.1 A Specific Factors Model

Consider an economy populated by three types of individuals: capitalists, landowners, and laborers. All individuals have similar preferences defined over two consumer goods (c_1, c_1). For simplicity, let the utility function be $u(c_1+c_2) = c_1+c_2$. With this assumption, the relative price of the two goods is fixed at unity (so that they are perfect substitutes).

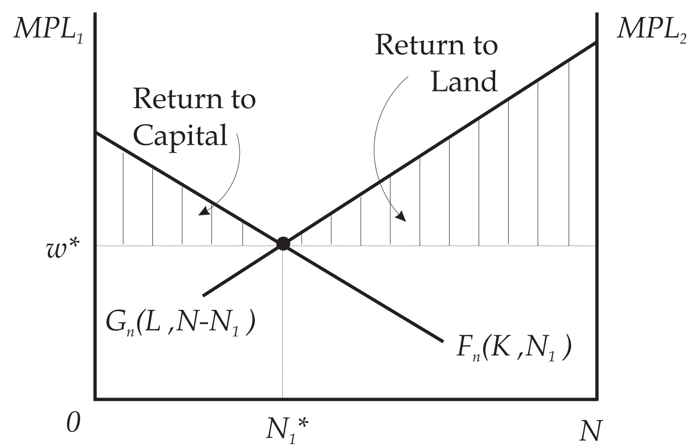
Capitalists are endowed with a fixed amount of physical capital K. Landowners are endowed with a fixed amount of land capital L. there are N laborers, each of whom are endowed with one unit of human capital. The two goods are produced according to the following technologies:

$$Y_1 = F(K, N_1);$$
$$Y_2 = G(L, N_2);$$

where N_j represents the number of workers employed in sector j. Assume that F and G are regular neoclassical production functions. Notice that Y_1 is produced with capital and labor, while Y_2 is produced with land and labor. In this model, capital and land are factors of production that are specific to a particular sector. Labor is used in the production of both goods (so you can think of labor as a general factor of production). Assume that labor is freely mobile across sectors.

Let F_n and G_n denote the marginal product of labor associated with each production function. Since F and G are neoclassical production functions, F_n and G_n are positive and declining in the labor input (i.e., there are diminishing returns to labor, given a fixed capital stock). In a competitive economy, labor is paid its marginal product. If labor is mobile across sectors, then the sectoral composition of employment will adjust to the point $F_n(K, N_1^*) = G_n(L, N_2^*)$, with $N_1^*+N_2^* = N$. Figure 10.4 displays the general equilibrium of this economy.

FIGURE 14.4
General Equilibrium
Specific Factors Model

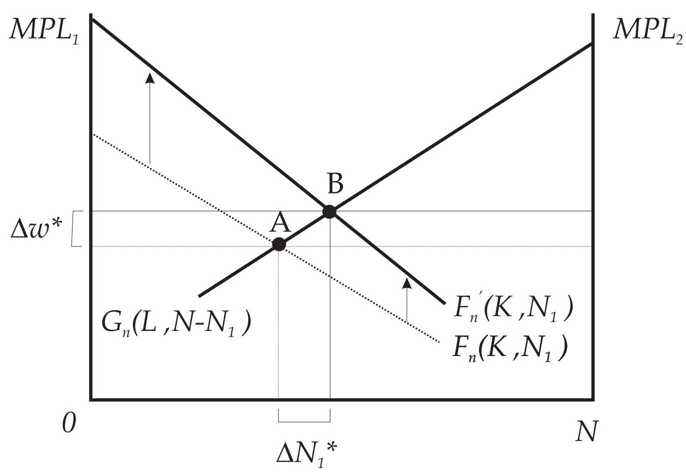


The triangular regions in Figure 14.4 represent the total returns to the specific factors. The rates of return accruing to each specific factor need not be equated, since these factors are, by assumption, immobile. The total return to labor is given by w^*N, with each worker earning the return w^*. In equilibrium, each worker is indifferent between which sector to work in, since each sector pays the same wage rate. To see why both sectors pay the same wage, suppose that they did not. For example, imagine that $w_1 > w_2$. Then workers in sector 2 would flock to sector 1, leading to an increase in N_1 (and a corresponding decline in N_2). But as N_1 expands, the marginal product of labor must fall in sector 1. Likewise, as N_2 contracts, the marginal product of labor in sector 2 must rise. In equilibrium, both sectors must pay the same wage.

Now, let us suppose that the capitalists of this economy realize that there is some better technology out there F' for producing the output Y_1. The effect of implementing this new technology is similar to the effect of a positive productivity shock studied in earlier chapters. The new technology has the effect of increasing the marginal product of labor at any given employment level N_1. Imagine that the new technology shifts the MPL_1 up, making it steeper, as in Figure 14.5.

Point A in Figure 14.5 depicts the initial equilibrium and point B depicts the new equilibrium. Notice that the new technology implemented in sector 1 has led to an increase in the wage in *both* sectors. The economic intuition for this is as follows. Since the new technology improves the marginal product of labor in sector 1, the demand for sector 1 workers increases. The increase in

FIGURE 14.5
A Sector-Specific Technology Shock



demand for labor puts upward pressure on the real wage, which attracts sector 2 workers to migrate to sector 1. Once again, in equilibrium, the wage must be equated across sectors so that any remaining workers are just indifferent between migrating or staying at home.

The new technology also increases the return to capital, since the new technology makes both capital and labor more productive. However, note that the increase in the economy-wide wage rate *reduces* the return to land. This is because the new technology does not increase the efficiency of production in sector 2. Yet, sector 2 landowners must pay their labor a higher wage, increasing their costs and hence lowering their profit.

The political economy implications of this simple model are rather straightforward. If landowners (in this example) wield enough political power, they may be able to 'block' the implementation of the superior technology. While per capita incomes will be lower as a result, *their* incomes will be spared the adverse consequences of a new technology that serves to reduce the value of their endowment (land, in this example).

- **Exercise 14.3.** Re-examine the specific factors model described above by assuming that the two fixed factors constitute 'high-skill' and 'low-skill' labor, with a freely mobile factor (capital) that is used in both sectors of the economy. Explain how a skill-biased technological advance may harm low-skill workers. Now imagine that low-skill workers are highly organized (represented by strong unions) and explain the pressure that politicians may face to 'regulate' the adoption of the new technology. What other ways may such workers be compensated?

14.3.2 Historical Evidence

It is important to realize that barriers emanating from special interests have always been present in all economies (from ancient to modern and rich to poor), so that any differences are really only a matter of degree. For example, as early as 1397, tailors in Cologne were forbidden to use machines that pressed pinheads. In 1561, the city council of Nuremberg, apparently influenced by the guild of red-metal turners, launched an attack on Hans Spaichl who had invented an improved slide rest lathe. The council first rewarded Spaichl for his invention, then began to harass him and made him promise not to sell his lathe outside his own craft, then offered to buy it from if he suppressed it, and finally threatened to imprison anyone who sold the lathe. The ribbon loom was invented in Danzig in 1579, but its inventor was reportedly secretly drowned by the orders of the city council. Twenty five years later, the ribbon loom was reinvented in the Netherlands (and so became known as the Dutch loom), although resistance there too was stiff. A century and a half later, John Kay, the inventor of the flying shuttle, was harassed by weavers. He eventually settled in France, where he refused to show his shuttle to weavers out of fear. In 1299, an edict was issued in Florence forbidding bankers to use Arabic numerals. In the fifteenth century, the scribes guild of Paris succeeded in delaying the introduction of the printing press in Paris by 20 years. In the sixteenth century, the great printers revolt in France was triggered by labor-saving innovations in the presses.

Another take on the special interest story pertains to case in which the government itself constitutes the special interest, as in the case of autocratic rulers. It seems as a general rule, weaker governments are able to exert less resistance to technological adoption. With some notable exceptions, autocratic rulers have tended to be hostile or indifferent to technological change. Since innovators are typically nonconformists and since technological change typically leads to disruption, the autocrat's instinctive desire for stability and suspicion of nonconformism could plausibly have outweighed the perceived gains to technological innovation. Thus, in both the Ming dynasty in China (1368–1644) and the Tokugawa regime in Japan (1600–1867) set the tone for inward-looking, conservative societies. Only when strong governments realized that technological backwardness itself constituted a threat to the regime (e.g., post 1867 Japan and modern day China) did they intervene directly to encourage technological change.

During the start of the Industrial Revolution in Britain, the political system strongly favored the winners over the losers. Perhaps this was because the British ruling class had most of its assets in real estate and agriculture which, if anything, benefited from technological progress in other areas (e.g., by increasing land rents). However, even in Britain, technological advances were met by stiff opposition. For example, in 1768, 500 sawyers assaulted a mechanical saw mill in London. Severe riots occurred in Lancashire in 1779, and there many instances of factories being burned. Between 1811 and 1816, the Midlands and the industrial counties were the site of the 'Luddite' riots, in which much damage was inflicted on machines. In 1826, hand-loom weavers in a few Lancashire towns rioted for three days. Many more episodes like these have been recorded.

But by and large, these attempts to prevent technological change in Britain were unsuccessful and only served to delay the inevitable. An important reason for this is to be found in how the government responded to attempts to halt technological progress. In 1769, Parliament passed a harsh law in which the wilful destruction of machinery was made a felony punishable by death. In 1779, the Lancashire riots were suppressed by the army. At this time, a resolution passed by the Preston justices of peace read: "The sole cause of the great riots was the new machines employed in cotton manufacture; the country notwithstanding has greatly benefited by their erection and destroying them in this country would only be the means of transferring them to another...to the great detriment of the trade of Britain."

The political barriers to efficiency manifest themselves in many ways, from trade restrictions and labor laws to regulatory red tape. For example, a recent World Bank report (Doing Business in 2004: Understanding Regulation) documents the following. It takes two days to register a business in Australia, but 203 days in Haiti. You pay nothing to start a business in Denmark, while in Cambodia you pay five times the country's average income and in Sierra Leone, you pay more than 13 times. In more than three dozen countries, including Hong Kong, Singapore and Thailand, there is no minimum on the capital required by someone wanting to start a business. In Syria, the minimum is 56 times the average income; in Ethiopia and Yemen, it's 17 times and in Mali, six. You can enforce a simple commercial contract in seven days in Tunisia and 39 days in the Netherlands, but in Guatemala it takes more than four years. The report makes it clear, however, that good regulation is not necessarily zero regulation. The report concludes that Hong Kong's economic success, Botswana's stellar growth performance and Hungary's smooth transition (from communism) have all been stimulated by a good regulatory environment. Presumably, a 'good' regulatory environment is one which allows individuals the freedom to contract while at the same time providing a legal environment that protects private property rights

While there are certainly many examples of special interests working against the implementation of better technology, our political economy story is not without shortcomings. In particular, special interest groups are busy at work in *all* societies. The key question then is why different societies confer more or less power to various special interests. Perhaps some societies, such as the United States, have erected institutions that are largely successful at mitigating the political influence of special interests. These institutions may have been erected at a time when a large part of the population shared similar interests (e.g., during the American revolution).

But even if new technologies have sectoral consequences for the economy, it is still not immediately clear why special interests should pose a problem for the way an economy functions. For example, in the context of the model developed above, why do individuals not hold a diversified portfolio of assets that would to some extent protect them from the risks associated with sector-specific shocks? In this way, individuals who are diversified can share in the gains of technological progress. Alternatively (and perhaps equivalently), why do the winners not compensate (bribe) the losers associated with a technological improvement? These and many other questions remain topics of current research.

14.4 Endogenous Growth Theory

To this point, we have assumed that technological progress occurs exogenously. The models developed to this point have concentrated on economic behavior *given* the nature of the technological frontier and how it evolves over time. But now it is time to think about what determines this frontier and its development over time. The large literature that has emerged recently to deal with this question is called *endogenous growth theory* and was spawned largely through the work of Paul Romer (1986, 1994).

While the Solow model is useful for thinking about the determinants of the *level* of long-run living standards, it is not capable (or designed to) explain the determinants of long-run growth *rates* (although the model is capable of explaining the short-run dynamics toward a steady state). In the absence of technological progress, growth in per capita income must eventually approach zero. The reason for this is to be found in the fact (or assumption) that capital accumulation is subject to *diminishing returns*.

But now let us think of knowledge itself as constituting a type of capital good. Unlike physical capital goods (or any other physical input), there are no obvious limitations to the expansion of knowledge. Romer has argued that the key feature of knowledge capital is its *nonrivalrous* nature. A good is said to be *rivalrous* if its use by one person excludes its use by someone else. For example, if I use a lawnmower to cut my lawn, this precludes anyone else from using my lawnmower to cut their lawn at the same time. Knowledge, on the other hand, is a different matter. For example, if I use a theorem to prove a particular result, this does not preclude anyone else from using the same theorem at the same time for their own productive purposes. Because of the nonrival nature of technology, there are no obvious diminishing returns from knowledge acquisition.

14.4.1 A Simple Model

Let z_t denote the stock of 'knowledge capital' available at date t, for $t = 1, 2, ..., \infty$. Output per capita is given by the production function $y_t = z_t f(k_t)$. For simplicity, let us assume a constant population and a constant stock of physical capital and normalize units such that $f(k) = 1$. In this case, we have $y_t = z_t$. What this says is that per capita output will grow *linearly* with the stock of knowledge. In other words, there are *constant returns to scale* in knowledge acquisition.

The economy is populated by two-period-lived overlapping generations, who has preferences defined over sequences of consumption $(c_t(j))$ and leisure (l),

$$U_t = \ln(c_t(1)) + v(l) + \beta \ln(c_{t+1}(2))$$

where $0 < \beta < 1$ is an exogenous time-preference parameter and v is an increasing and strictly concave function. Here, $c_t(j)$ represents the consumption of an individual in period t in the j^{th} period of life. For these preferences, the marginal rate of substitution between current leisure (l) and future consumption (c_{t+1}) is given by:

$$MRS(l, c_{t+1}(2)) = \frac{c_{t+1}(2)v'(l)}{\beta}.$$

As well, the marginal rate of substitution between consumption at two different points in time is given by:

$$MRS(c_t(1), c_{t+1}(2)) = \frac{1}{\beta}\frac{c_{t+1}(2)}{c_t(1)}.$$

Young Individuals are endowed with two units of time and old individuals are endowed with one unit of time. One unit of this time is used (exogenously) in production, which generates $y_t = z_t$ units of output. In a competitive labor market, z_t would also represent the equilibrium real wage. Output is nonstorable (the physical capital stock cannot be augmented), so that $c_t(1) = y_t = z_t$ and $c_{t+1}(2) = y_{t+1} = z_{t+1}$. The remaining unit of time for young individuals can be used in one of two activities: leisure (l) or learning effort (e). Thus, individuals are faced with the time constraint:

$$e + l = 1. \tag{14.9}$$

Learning effort can be thought of the time spent in R&D activities. While diverting time away from leisure is costly, the benefit is that learning effort augments the future stock of knowledge capital. We can model this assumption in the following way:

$$z_{t+1} = (1 + e)z_t. \tag{14.10}$$

Observe that $e = 0$ implies that $z_{t+1} = z_t$ and that $e > 0$ implies $z_{t+1} > z_t$. In fact, e represents the rate of growth of knowledge (and hence the rate of growth of per capita GDP).

Since $c_t(1) = z_t$ and $c_{t+1}(2) = z_{t+1}$, we can combine equations (14.10) and (14.9) to form a relationship that describes the trade off between current leisure and future consumption:

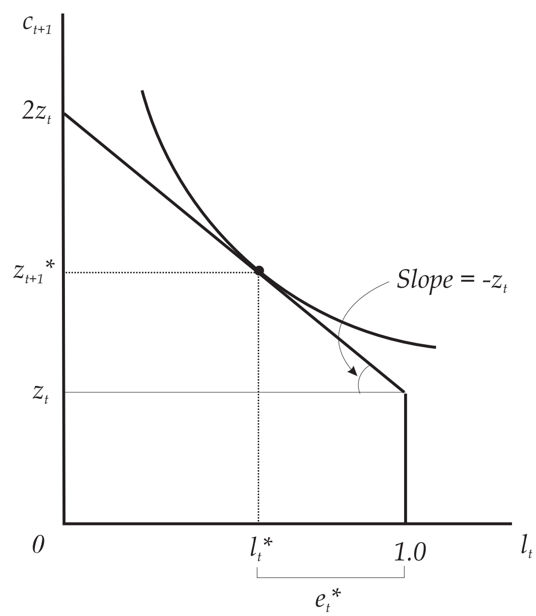
$$c_{t+1}(2) = (2 - l)z_t. \tag{14.11}$$

This constraint tells us that if $l = 1$ (so that $e = 0$), then $c_{t+1}(2) = z_t = c_t(1)$ (consumption will remain the same). On the other hand, if $l = 0$ (so that $e = 1$), then $c_{t+1}(2) = 2z_t = 2c_t(1)$ (consumption will double from this period to the next). In general, individuals will choose some intermediate level of l (and hence e) such that the marginal cost and benefit of learning effort is just equated. In other words,

$$\frac{c_{t+1}(2)v'(l)}{\beta} = z_t. \tag{14.12}$$

Conditions (14.12) and (14.11) are depicted in Figure 14.6.

FIGURE 14.6
Equilibrium Growth Rate



Notice that since $c_{t+1}(2) = z_{t+1}$, we can rewrite condition (14.12) as:

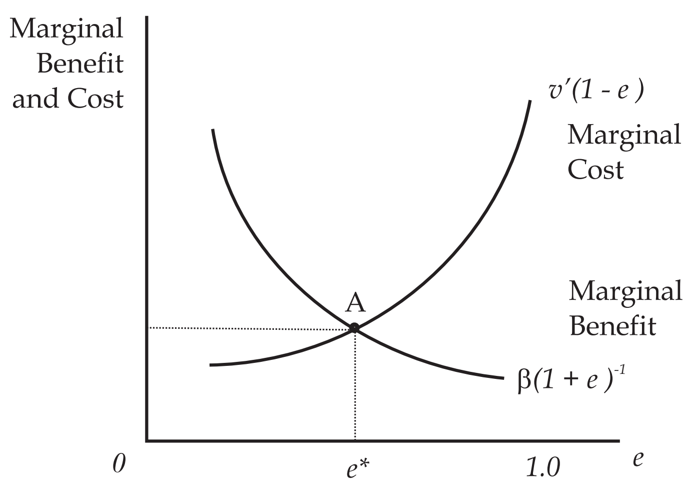
$$v'(l) = \beta \frac{z_t}{z_{t+1}} = \beta \left(\frac{1}{1+e} \right),$$

or,

$$v'(1 - e^*) = \beta \left(\frac{1}{1+e^*} \right). \tag{14.13}$$

Equation (14.13) is one equation in the one unknown, e^*. Thus, this condition can be used to solve for the equilibrium growth rate e^*. The left hand side of equation (14.13) can be thought of as the marginal utility cost of learning effort. Since v is strictly increasing and concave in leisure, increasing e (reducing leisure) increases the marginal cost of learning effort. The right hand side of equation (14.13) can be thought of as the marginal utility benefit of learning effort. An increase in learning effort increases future consumption, but since u is concave, the marginal benefit of this extra consumption falls with the level of consumption. Figure 14.7 displays the solution in (14.13) with a diagram. The equilibrium steady state growth rate is determined by the condition that the marginal cost of learning effort is just equated with the marginal benefit; i.e., point A in Figure 14.7.

FIGURE 14.7
Steady State Growth Rate



Note that for the specification of preferences that we have assumed, the equilibrium growth rate does not depend on z_t.

- **Exercise 14.4.** Using a diagram similar to Figure 14.6, show how it is possible for an increase in z_t not to affect the equilibrium growth rate. Provide some economic intuition.

With e^* determined in this manner, the equilibrium growth rate in per capita GDP is given by $(y_{t+1}/y_t) = (1 + e^*)$. Note that long-run growth is endogenous in this model because growth is not assumed (instead, we have derived this property from a deeper set of assumptions). In particular, note that zero growth is feasible (for example, by setting $e^* = 0$). In general, however, individuals will find it in their interest to choose some positive level of e^*.

Finally, note that we can derive an expression for the equilibrium real rate of interest in this economy. Since the marginal rate of substitution between time-dated consumption is given by $MRS = c_{t+1}/(\beta c_t)$, we can use what we learned in Chapter 6 by noting that the desired consumption profile must satisfy:

$$\frac{c_{t+1}}{\beta c_t} = R_t,$$

where R_t is the (gross) real rate of interest (which is earned on or paid for risk-free private debt). Because this is a closed endowment economy, we know that the interest rate must adjust to ensure that desired national savings is equal to zero. In other words, it must be the case that $c_t^* = z_t$ for every date t. Thus, it follows that:

$$R^* = \frac{1}{\beta}(1 + e^*). \tag{14.14}$$

14.4.2 Initial Conditions and Nonconvergence

Now, let us consider two economies that are identical to each other in every way except for an initial condition z_1. For example, suppose that $z_1^A > z_1^B$, so that economy A is initially richer than economy B. Furthermore, assume that each economy remains closed to international trade and knowledge flows (so that economy B cannot simply adopt z_1^A). Then our model predicts that these two economies will grow at exactly the same rate (see Exercise 14.4). In other words, economy A will forever be richer than economy B; i.e., the levels of per capita GDP will never converge. Note that this result is very different than the prediction offered by the Solow model, which suggests a long-run convergence of growth rates (to zero).

Some economists have proposed this type of explanation for the apparent lack of convergence in per capita GDP across many countries. In Figure 10.1, for example, we see that many countries (in this sample) are growing roughly at the same rate as the world's technological leader (the United States).[3]

As with many models, there is probably an element of truth to this story. But the theory also has its challenges. For example, I noted in the introduction that according to our best estimates, the 'initial conditions' around the world c. 1800 were not that different. On the other hand, even very small differences in initial conditions can manifest themselves as very large differences in levels over an extended period of time. But then again, what might explain the fact that some countries have managed to grow much faster than the U.S. for prolonged periods of time? Perhaps this phenomenon was due to an international flow of knowledge

[3] Remember that since Figure 10.1 plots a country's per capita GDP relative to the U.S., a 'flat' profile indicates that the country is growing at the same rate as the U.S.

that allowed some countries to 'catch up' to U.S. living standards. A prime example of this may be post war Japan, which very quickly made widespread use of existing U.S. technology. Similarly, growth slowdowns may be explained by legal restrictions that prevent the importation of new technologies. At the end of the day, the million dollar question remains: Why do lesser developed countries not do more to encourage the importation of superior technologies? In other words, why don't countries simply imitate the world's technological leaders?

14.5 References

1. Bairoch, P. (1993). *Economics and World History: Myths and Paradoxes*, New York: Harvester-Wheatsheaf.

2. Parente, Stephen L. and Edward C. Prescott (1999). "Barriers to Riches," *Third Walras-Pareto Lecture*, University of Lausanne.

3. Pomeranz, K. (1998). "East Asia, Europe, and the Industrial Revolution," Unpublished Ph.D. Thesis, University of California (Irvine).

4. Romer, Paul M. (1986). "Increasing Returns and Long-Run Growth," *Journal of Political Economy*, 94: 1002–1037.

5. Romer, Paul M. (1994). "The Origins of Endogenous Growth," *Journal of Economic Perspectives*, 8: 3–22.

6. Solow, Robert M. (1956). "A Contribution to the Theory of Economic Growth," *Quarterly Journal of Economics*, 70: 65–94.