Macroeconomic Theory and Policy

David Andolfatto
Department of Economics
Simon Fraser University
2006

\[1\] © This text is not for sale, but may be distributed freely.
Contents

Preface

Part 1: Basic Theory

Chapter 1: The Gross Domestic Product
Chapter 2: Output and Employment
Chapter 3: Government Spending and Taxation
Chapter 4: Consumption and Saving
Chapter 5: Government Deficits
Chapter 6: Capital and Investment
Chapter 7: Unemployment

Part 2: Monetary Theory

Chapter 8: Money
Chapter 9: Inflation
Chapter 10: Money and Capital
Chapter 11: International Monetary Systems

Part 3: Growth and Development Theory

Chapter 12: Early Economic Development
Chapter 13: Modern Economic Development
Preface

The field of macroeconomic theory has evolved rapidly over the last quarter century. This evolution has manifested itself primarily in terms of methodological developments, rather than in the set of questions that interest macroeconomists. These methodological developments entail the use of language and tools that are based on microeconomic theory (including game theory). Modern cohorts of PhD students are taught how to interpret macroeconomic phenomena in terms of model economies consisting of economic actors with explicit goals facing explicit constraints. Aggregate outcomes (model predictions) must satisfy a specified consistency (equilibrium) requirement. This methodology is in stark contrast to earlier methods that rely primarily on aggregate reduced-form equations to describe how an economy functions. Unfortunately (from my perspective), most introductory and intermediate macroeconomic textbooks today continue to employ these earlier (and outdated) methods. It is time to move on. This text is designed to help.

The main challenge in teaching macroeconomics to undergraduates using the modern methodology has to be the added technical (primarily mathematical) requirements that the approach demands. The frontier of the discipline utilizes models that are so complicated, they can only be solved using numerical methods. But while one cannot really do justice to the richness of the modern approach, I believe that one can convey many basic lessons and insights by employing a sequence of simplified models that build on each other. Much, if not all, of this basic intuition can be exposited by way of simple diagrams (and high-school algebra). This is what I try to do. You can judge for yourself whether you think I am successful in this endeavor.

I would like to comment on a criticism that one often hears regarding the modern approach; in particular, its apparent reliance on the ‘representative agent’ hypothesis. As it turns out, I do make extensive (but not exclusive) use of this simplifying assumption. One should be aware, however, that the modern approach in no way depends on the representative agent formulation (there is nothing that prevents one from introducing as much heterogeneity that is desired). The representative agent hypothesis is used nowadays primarily as a pedagogical device. While some conclusions are no doubt sensitive to the assumption, there are many that are not (and it is these latter conclusions that deserve emphasis). In any case, I find it ironic that the criticism is most often leveled by those who prefer the older methodology; which, by its very nature, is typically cast in terms of a representative agent (e.g., in the form of an aggregate consumption function that makes no reference to individual differences).

I like the modern approach for two reasons. First, it highlights the interaction between incentives and constraints that govern decision-making. Economics is about decision-making and macroeconomics is concerned with explaining how
individual decisions interact to generate aggregate outcomes. This decision-making process is almost entirely neglected in the older approach, which simply describes (rather than explains) why people behave the way they do. Second, because the modern approach insists on modeling preferences explicitly, there is a straightforward and entirely natural way to evaluate the welfare implications of government policy interventions. In contrast, the older approach must rely exclusively on the modeler’s own *ad hoc* welfare function to make similar statements. For example, it is often simply assumed that higher income, or lower unemployment, or a stabilized business cycle must be welfare improving. As it turns out, these assumptions have little or no theoretical foundation. A qualified professional may understand the circumstances under which such an assumptions are justified; but a student trained solely in older methods is not likely to appreciate such subtleties (which possibly accounts for why we see so many crazy policies implemented).

Alright then, enough blathering. Either you like the text or you don’t. Keep in mind that this is still work in progress (I have been revising these notes yearly). At least you can’t complain about the price! The manuscript can be downloaded for free at:

www.sfu.ca/~dandolfa/macro2006.pdf

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 Jiang. 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
CHAPTER 1

The Gross Domestic Product

1. Introduction

The Gross Domestic Product (GDP) is an economic statistic that one hears quoted frequently in the news and elsewhere. But what exactly is this GDP thing supposed to measure? And why should anyone care about whether it is measured at all?

Most people have at least a vague idea that the GDP represents some measure of ‘economic performance.’ One often hears, for example, that a country with a higher GDP is performing better than one with a lower GDP; or that a rapidly growing GDP is better than a stable, or declining GDP. This idea of the GDP as a measure of economic performance is held so widely and (at times) accepted so uncritically, that on these grounds alone, it probably deserves closer scrutiny.

Before we can talk sensibly about GDP and why it might matter, we should have a clear understanding of how the term is defined and measured. Most countries in the world have a government agency (or agencies) responsible for collecting and aggregating measures of economic activity. You can find a list of these agencies at the following website (the United Nations Statistics Division):


In Canada, our national statistical agency is called Statistics Canada.\(^2\) Among other things, Statistics Canada maintains a system of national **Income and Expenditure Accounts** (IEA). The following quote, taken from the Statistics Canada webpage, sums up their own naive view of the world:

The Income and Expenditure Accounts are the centre of macroeconomic analysis and policy-making in Canada. They are a means by which Canadians can view and assess the performance of the national economy. The accounts provide both a planning framework for governments and a report card on the results of the plans that governments carry out. At the core of the Income and Expenditure Accounts (IEA) is the concept of Gross domestic product (GDP) and its components.

\(^2\)See: http://www.statcan.ca/
The statement above makes clear that GDP is considered a core concept. So let’s take some time to investigate its measurement and potential usefulness.

2. Definition of GDP

Here is a standard definition of GDP:

\[ \text{GDP: The total value of final goods and services produced in the domestic economy over some given period of time.} \]

From this definition, we gather that the GDP represents some measure of the level of production in an economy. For this reason, the GDP is commonly referred to as output. Keep in mind that output constitutes a flow of goods and services. That is, it represents the value of what is produced over some given interval of time (e.g., a month, a quarter, or a year). Food, clothing, and shelter services produced over the course of a year all contribute to an economy’s annual GDP.

Let us now examine the definition of GDP more carefully. Note first of all that the GDP measures the ‘value’ of output. We will discuss the concept of ‘value’ in some detail later on; but for now, assume that value is measured in units of ‘dollars’ (feel free to substitute your favorite currency). When output is measured in units of money, it is referred to as the nominal GDP (at current prices).

Output takes the form of goods and services. What is the difference between a good and a service? A good is an object that can be held as inventory; while a service is an object that cannot be stored. Think of the difference between an orange and a haircut. Note that any good is likely valued only to the extent that it yields (or is expected to yield) a service flow; as when I consume that orange, for example.

Next, note that the definition above makes reference to final goods and services. A final good is to be distinguished from an intermediate good. An intermediate good is an object that is produced and utilized as a input toward the production of some other good or service within the time period of consideration.

An example may help clarify. Imagine that last year, an economy produced $200 of vegetables, $150 of fertilizer, $100 of bread, and $50 of flour. Imagine further that all of the fertilizer was used in the production of vegetables and all of the flour was used in the production of bread. One might be tempted to conclude that the annual GDP for this economy is $500, but this would be wrong. In fact, the GDP is equal to $200 + $100 = $300; that is, the total value of the final goods produced (bread and vegetables). The value of the intermediate goods is excluded from this calculation because their value is already embedded in the value of the final goods. That is, when you pay $2.00 for a bundle of
carrots at the supermarket, $1.50 represents the value of the fertilizer that was used to grow these carrots.

The example above suggests an alternative (but equivalent) definition of GDP. Define value-added as the value of a good or service net of the cost of any intermediate inputs used to produce it. Then one can define the GDP as the total valued-added. In the context of the example above, we have $150 of fertilizer and $50 of flour (both of which use no intermediate inputs), together with the value added in the production of vegetables ($50) and bread ($50); the sum of which is $300.

Moving along, observe that the definition above makes reference to the domestic product. The domestic product is to be distinguished from the national product; otherwise known as the Gross National Product (GNP). The difference is as follows. The GDP measures the value of output produced within the borders of a domestic economy, whether or not this production takes place with foreign-owned factors of production. The GNP, on the other hand, measures the value of output produced by the factors of production owned by the ‘citizens’ of a domestic economy, whether or not these factors of production reside on domestic soil or not. For countries like Canada and the United States, the difference between GDP and GNP is relatively small. For countries like Turkey and Mexico, on the other hand—with many citizens living and working abroad—the discrepancy between GDP and GNP can be significant.

Finally, consider the term gross in the definition of GDP. Here, the gross domestic product is to be distinguished from the net domestic product (NDP). The NDP is defined as the GDP net of capital consumption. Capital consumption simply refers to the value of capital that is consumed (i.e., destroyed or depreciated) in the act of production.

A case could be made that the NDP is a better measure of actual production. For example, if construction workers destroy $20,000 worth of equipment in the process of building a $100,000 house, most people would probably agree that $80,000 constitutes a better measure of the value added to the economy. Environmentalists are particularly fond of an NDP measure modified to include ‘environmental degradation’ and ‘resource depletion’ as components of capital depreciation.

**Exercise 1.1:** Consider the example above, of an economy that produces $200 in vegetables, $150 in fertilizer, $100 in bread, and $50 in flour over the course of one year. As before, assume that the entire amount of fertilizer and flour is consumed in the process of producing vegetables and bread. But imagine now that the fertilizer and flour were not produced this year; that is, suppose that they were produced last year and brought over into this year as inventory (capital). Does the fertilizer and flour in this example fit the definition of an intermediate good? Compute the GDP and NDP for this economy.
3. Consumption vs. Investment

Consider all the millions of goods and services produced in an economy over the course of some time interval. Economists have found it useful to divide this vast and heterogeneous flow of output into two categories: consumption goods (and services); and investment goods (and services). What are the distinguishing characteristics of these two types of output and why is it useful to make such a distinction?

Consumption represents that part of the output flow that is consumed (i.e., destroyed) for the purpose of augmenting current material living standards. By ‘current,’ I mean over the course of a given time-interval, like one month or one year. Investment represents that part of the output flow that is destined to augment the future production of output (and ultimately, future material living standards). Investment is sometimes also referred to as the production of new capital goods and services. Note that new capital goods are to be distinguished from old capital goods—or, the stock of existing capital—which is presumably employed in the production of current output (along with other factors of production).

Consumption: That branch of the output flow that is consumed (destroyed) for the purpose of augmenting current material living standards.

One should keep in mind that the distinction between consumption and investment is not always so clear-cut; and in particular, the distinction may depend on the time-interval under consideration. Suppose, for example, that I hire the kid next door to mow my lawn one sunny afternoon. The kid’s labor (together with his lawnmower capital) is used to produce ‘lawn-enhancement services.’ I pay the kid $10 and let him have a sip of my beer (when his mother isn’t looking). Now, this $10 in lawn-enhancement services—does it constitute output in the form of consumption or investment?

The answer depends on the time-interval under consideration. Imagine that the benefit I derive from the freshly-cut lawn lasts for one week (the lawn needs to be mowed again after this period of time). If the time-interval under consideration is one day, then one could well argue that the lawn-enhancement service constituted a form of investment that generated a flow of consumption for several periods (i.e., days). On the other hand, if the time-interval under consideration is one week (or more), then one might argue that the lawn-enhancement service simply constituted consumption (as the output depreciates fully after one week).

Thus, one way to distinguish between consumption and investment is to fix the time-interval under consideration and ask whether newly-produced output is expected to last longer than this time interval. If the answer is no, then the output constitutes consumption. If the answer is yes, then the output
constitutes a form of investment that augments the stock of existing capital (with capital generating a flow of future services).

**Investment:** That branch of the output flow that augments the existing stock of capital.

The most common time-intervals used in macroeconomic analysis are one quarter (3 months) and one year. With these time-intervals, a large class of goods and services can be clearly categorized as either consumption or investment. The construction of a new house, a new piece of machinery, a new car, for example, would seem to constitute new capital goods (additions to the existing stock of capital). The same could be said of many medical procedures (from hernia operations to breast implants) and education services (to the extent that students can be expected to hold on to what they have learned beyond the final exam). In contrast, the production of perishable food products, transportation services, haircuts, shelter services, etc., would seem to constitute consumption. However, there remain other forms of output that are not so easily classified; for these objects, a judgement call must be made.

It is important to keep in mind that the term ‘investment’ as it is used here differs from the way it is commonly used in everyday language. Imagine, for example, that you are currently living in a rented apartment but decide to purchase a home. Most people would regard this purchase as an ‘investment’ in real estate. But whether this purchase is counted as investment in the IEA depends on whether the home you purchased is an old home or a new home. An old home is considered to be part of the existing (residential) capital stock. The purchase of an old home simply represents a change of ownership in the existing stock of capital and hence is not counted as investment for the economy as a whole. A new home, on the other hand, represents a new addition to the existing stock of residential capital; a new home is counted as investment for the economy as a whole.

The IEA definition of investment generally differs from the way people commonly understand the term.

Why is the distinction between consumption and investment important? This distinction is important because the manner in which an economy divides its output flow across consumption and investment ultimately determines the ‘long-run’ living standards of its inhabitants. Consider, for example, an economy consisting of farmers producing perishable food products year after year at some given level. This level of production determines living standards now and off into the indefinite future. Suppose now that some of these farmers become construction workers employed in the production of greenhouses and irrigation
systems. This diversion of labor necessarily entails a temporary decline in food production (and hence current living standards) as the new capital that is constructed takes time to be productive. But in the ‘long-run,’ the new capital has the effect of enhancing agricultural output and hence future living standards beyond the initial level. In this way, investment entails a sacrifice of current consumption in exchange for higher levels of future consumption.

On the Concept and Measurement of Capital

The term capital appears to mean different things to different people. To an economist, capital refers to a durable factor of production and inventory. It does not, in particular, refer to financial assets, which simply represent claims to future objects.

The most obvious form of capital is what is called physical capital. Examples of physical capital include objects like machinery, computers, buildings, land, automobiles, highways, sewage systems, and inventories of goods. Even physical objects such as these are difficult to measure. For example, in measuring the stock of residential capital, is it appropriate to count the number of houses and apartments, or the square footage of living space? And is a 3000 square foot home made of brick the same thing as a 3000 square foot home made out of rice paper? Furthermore, how does one add together a house and a printing press to arrive at an aggregate measure of physical capital?

In principle, perhaps the only way to measure capital consistently is by market value. In this way, we could say that a brick home valued at $400,000 and a printing press valued at $100,000 together make up $500,000 worth of capital. The ‘problem’ with this approach is that the value of capital may vary even without any change in its physical quantity. On the other hand, perhaps it makes sense to think of more valuable capital as constituting more ‘effective’ capital.\(^3\)

In practice, it appears that the aggregate stock of physical capital is measured using a ‘perpetual inventory method.’ The way this is done is to take the investment flow (measured at market value) in each of various asset classes, applying a constant depreciation rate (that varies with asset class) and then adding the results across investment years and asset classes. A major problem with this method is that it values the existing capital stock at book value (historical market value of past investment) instead of current market value. Presumably, this is done because it is difficult to ascertain the current market value of all forms of capital.

Unfortunately, the measurement of capital problem is in fact much worse than this. The reason for this lies in the fact that physical capital is not the only—or even most important—form of capital in an economy. One could reason

by analogy that every human being is a kind of durable machine. In a very real
sense, each person can be thought of as an owner-operator of human capital
that generates a (potential) stream of labor services over time.

Like a machine or a house, we need to be maintained and (at times) re-
paired. In the IEA, maintenance and repair is considered a form of investment.
In contrast, all personal expenditures on food, clothing, shelter and medical ser-
vices are treated as consumption. Does this make sense? This is to say nothing
about the investments that people make to improve their skill (again, education
is treated as a form of consumption instead of investment), or the investments
that parents make in raising their children.

What is the value of human capital? In theory, the market value of a person’s
human capital is the present value of one’s lifetime wage stream net of the
present value of any direct investments in human capital (the same principle
holds for the valuation of any form of capital). In practice, direct claims on
human capital are rarely exchanged in markets, making the value of human
capital more difficult (but not impossible) to estimate.

Another important form of ‘intangible’ capital exists in the form of (disem-
body) technology. The modifier ‘disembodied’ here is used in reference to
‘knowledge’ or ‘technological know-how’ that exists separately from what may
be embodied in either physical or human capital. Some examples that come to
mind here may include things like ‘organizational capital’ (the way production
activities and distribution networks are organized, or other ‘best-practice’ tech-
niques), or even the introduction of new products (e.g., the sudden availability
of computers is not the same thing as having more factory space).

As you may have guessed, the value of these ‘intangible’ objects is often
difficult to measure. Nevertheless, this does not diminish their potential im-
portance. Firms can and do spend significant resources toward ‘figuring things
out.’ The most obvious example is R&D expenditure. R&D spending is clearly
a form of investment, even if the value of what is produced by such spending
is difficult to measure. But for some strange reason, R&D spending (and other
forms of investment in ‘intangible’ capital) is not counted as investment in the
IEA. It is, however, counted as part of the GDP. Implicitly then, R&D spending
is counted as form of consumption.

The upshot of all this is that the IEA essentially ignores human and intan-
gible capital, so that care must be exercised in interpreting the measured stock
of capital as the true stock of productive capital. Likewise, one must be careful
in interpreting measured investment as reflecting the true level of investment in
an economy (e.g., see Figure 1.1).
Big Companies Go **Intangible**

Companies are putting more emphasis on R&D and less on capital investment. Since 2000, the "intangibility index"—the ratio of R&D to capital spending, multiplied by 100—has risen for 9 of the 10 biggest U.S. companies that report R&D.

<table>
<thead>
<tr>
<th>COMPANY</th>
<th>INTANGIBILITY INDEX* 2000</th>
<th>LATEST**</th>
</tr>
</thead>
<tbody>
<tr>
<td>EXXONMOBIL</td>
<td>5.1</td>
<td>4.4</td>
</tr>
<tr>
<td>GE***</td>
<td>73.6</td>
<td>100.7</td>
</tr>
<tr>
<td>MICROSOFT</td>
<td>429.1</td>
<td>761.6</td>
</tr>
<tr>
<td>PROCTER &amp; GAMBLE</td>
<td>62.9</td>
<td>89.0</td>
</tr>
<tr>
<td>PFIZER</td>
<td>211.0</td>
<td>295.4</td>
</tr>
<tr>
<td>JOHNSON &amp; JOHNSON</td>
<td>183.8</td>
<td>239.2</td>
</tr>
<tr>
<td>ALTRIA</td>
<td>32.0</td>
<td>42.3</td>
</tr>
<tr>
<td>CHEVRONTEXACO</td>
<td>2.2</td>
<td>2.9</td>
</tr>
<tr>
<td>INTEL</td>
<td>58.4</td>
<td>88.4</td>
</tr>
<tr>
<td>IBM</td>
<td>95.6</td>
<td>129.9</td>
</tr>
<tr>
<td><strong>ALL 10</strong></td>
<td>56.8</td>
<td>79.1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>OVERALL</th>
<th>PERCENTAGE CHANGE 2000- LATEST**</th>
</tr>
</thead>
<tbody>
<tr>
<td>R&amp;D SPENDING</td>
<td>+42.1%</td>
</tr>
<tr>
<td>CAPITAL SPENDING</td>
<td>+2.1</td>
</tr>
</tbody>
</table>

* Capital spending for oil companies includes expenditures for exploration as well.
** Latest year for which R&D and capital spending are both available.
*** Excluding GE Capital Services

Data: Company reports, BusinessWeek
4. How the GDP is Calculated

As the son of Italian immigrants, I have had the pleasure of participating in the process of sausage-making from the ground up (including the initial slaughter). Believe me, this is an experience you can definitely afford to miss (I have never looked at my breakfast sausage in quite the same way ever since). And while I have never worked at a statistical agency, I have friends who have. From what I can gather, they now view ‘official statistics’ much in the same way I view my breakfast sausage.

The IEA report two measures of GDP, both of which should add up to the same number in theory (but in practice differ by a relatively small number called a ‘statistical discrepancy’). These two measures are based on two different approaches: an income approach, and an expenditure approach. Below, I discuss each approach in turn.

The Expenditure Approach

The expenditure approach to computing the GDP relies on the following fact: Everything that is produced must also be purchased. At first blush, this might seem like a strange thing to assert. What if, for example, a manufacturer produces an automobile that is not delivered to a dealer? In this case, the newly produced automobile is treated as a purchase of inventory by the manufacturer. The value of this inventory investment is based on the market price of the automobile (i.e., the market value of similar automobiles that are sold on the market).

Thus, the expenditure approach calculates GDP as the total spending on all domestically-produced final goods and services. The GDP calculated in this manner is sometimes referred to as the Gross Domestic Expenditure (GDE). Mathematically, this is done as follows. Let $x_{it}$ denote the quantity of good (or service) $i$ that is sold at date $t$ at market-price $p_{it}$. Then the value of this expenditure (measured in dollars) is simply $p_{it}x_{it}$. If there are $Q$ such goods and services, then the GDE is given by:

$$GDE_t \equiv \sum_{i=1}^{Q} p_{it}x_{it}. \quad (1)$$

Again, note that expenditures on intermediate goods and services are not included in this calculation (why not?).

---

4The notation $x \equiv y$ means that $x$ is equivalent (or by definition) equal to $y$. Note that
Actually, this is not quite how it’s done. The way it is done in practice is to first compute total spending on all (newly produced) final goods and services, whether or not they are domestically produced. Of course, this figure will include expenditures on imports, which are goods and services that are not produced domestically. To arrive at the GDE then, one must subtract from this figure the total value of all imported goods and services.

Largely as the result of an historical accident, the national income and product accounts organize the expenditure components of the GDP into four broad categories that depend on the sector in which the expenditure is undertaken. This classification is somewhat arbitrary in that there is no unique way in which to define ‘sector.’ Nevertheless, the way this is done in practice is to define four sectors as follows: [1] the household sector; [2] the business sector; [3] the government sector; and [4] the foreign sector. Sometimes, [1] and [2] are combined to form the domestic private sector. The government sector includes all levels of government (i.e., federal, provincial, state, local, etc.). The foreign sector includes both foreign private and government agencies.

Having defined sectors in this way, let $H_t$ denote household spending; let $B_t$ denote business spending; let $G_t$ denote government spending; and let $X_t$ denote foreign sector spending (on domestically-produced goods and services). As the spending on $C_t$, $I_t$ and $G_t$ includes imports, we have to subtract off the value of these imports $M_t$ to calculate spending on domestically-produced output. Using these expenditure categories, the GDE may equivalently be calculated as:

$$GDE_t \equiv H_t + B_t + G_t + X_t - M_t.$$  \hspace{1cm} (2)

Since this is probably not your first macro class, you’ve likely seen something similar to (2) before. It doesn’t quite look right though, does it? This is because every macroeconomics textbook in existence (that I am aware of) uses slightly different notation; and instead writes (2) in the following way:

$$GDE_t \equiv C_t + I_t + G_t + X_t - M_t.$$  \hspace{1cm} (3)

Obviously, (2) and (3) are equivalent if we define $C_t \equiv H_t$ and $I_t \equiv B_t$. There is nothing wrong in using whatever notation we wish, as long as the notation does not detract from clear thinking.

Unfortunately, the widely-used notation in (3) does at times appear to detract from clear thinking. Let me explain. Get your hands on any macro text currently on the market and flip to the section on national income accounting. Now look for the expenditure identity (3). In the discussion that surrounds this identity, you will invariably find statements asserting that $C_t$ denotes consumption expenditure and $I_t$ denotes investment expenditure. These statements are misleading (a product of bad notation) because $C_t$ in fact represents household consumption while $I_t$ represents investment. For example, if $x$ denotes ‘supply’ and $y$ denotes ‘demand,’ then it is not true that $x \equiv y$. However, it is true that $x = y$ at a market-clearing price.

\[\text{this is not the same as stating } x = y. \text{ The latter is an equation while the former is an identity.}\]

\[\text{For example, if } x \text{ denotes ‘supply’ and } y \text{ denotes ‘demand,’ then it is not true that } x \equiv y. \text{ However, it is true that } x = y \text{ at a market-clearing price.}\]
sector spending on both consumption and investment (in the form of durables and human capital investments) and \( I_t \) represents only one component of investment (i.e., each of \( C_t, G_t \) and \( X_t \) also include expenditures on investment goods and services).

Table 1.1 provides the expenditure components of GDP based on (2) for Canada. A number of observations are in order here. First, as remarked earlier, note that one should refrain from interpreting the subtotal \( C \) as consumption. That is, a significant component of household expenditures on goods are in the form of durable and semi-durable goods. Furthermore, one can guess that a significant component of the purchase of services is also in the form of investment services (broadly defined). Second, again as remarked earlier, note that the IEA only appears to count business sector spending \( I \) in the form of additions to physical capital (and not any investments in intangible capital). Third, note that a portion of government purchases \( G \) is in the form of new capital goods. It is probably the case, however, that some of the spending categorized as ‘current’ goods and services might be better labelled as investment. Fourth, note that expenditures on exports \( X \) generally consist of both consumption and investment goods and services.

Table 1.1 Gross Domestic Product: Expenditure-Based
Canada 2005 (millions of dollars)

<table>
<thead>
<tr>
<th>Sector</th>
<th>Subtotal (in millions of dollars)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Household Sector</strong></td>
<td></td>
</tr>
<tr>
<td>Durable and semi-durable goods</td>
<td>164,815</td>
</tr>
<tr>
<td>Non-durable goods</td>
<td>189,213</td>
</tr>
<tr>
<td>Services</td>
<td>407,934</td>
</tr>
<tr>
<td>Subtotal (( C ))</td>
<td>761,962</td>
</tr>
<tr>
<td><strong>Business Sector</strong></td>
<td></td>
</tr>
<tr>
<td>Residential structures</td>
<td>89,595</td>
</tr>
<tr>
<td>Non-residential structures</td>
<td>63,938</td>
</tr>
<tr>
<td>Machinery and equipment</td>
<td>91,354</td>
</tr>
<tr>
<td>Inventory investment</td>
<td>9,469</td>
</tr>
<tr>
<td>Subtotal (( I ))</td>
<td>254,356</td>
</tr>
<tr>
<td><strong>Government Sector</strong></td>
<td></td>
</tr>
<tr>
<td>Current goods and services</td>
<td>262,369</td>
</tr>
<tr>
<td>Investment</td>
<td>35,156</td>
</tr>
<tr>
<td>Subtotal (( G ))</td>
<td>297,525</td>
</tr>
<tr>
<td><strong>Foreign Sector</strong></td>
<td></td>
</tr>
<tr>
<td>Exports of goods</td>
<td>443,401</td>
</tr>
<tr>
<td>Exports of services</td>
<td>64,855</td>
</tr>
<tr>
<td>Subtotal (( X ))</td>
<td>518,256</td>
</tr>
<tr>
<td><strong>Imports of goods</strong></td>
<td>386,749</td>
</tr>
<tr>
<td><strong>Imports of services</strong></td>
<td>77,281</td>
</tr>
<tr>
<td><strong>Deduct: Subtotal (( M ))</strong></td>
<td>464,030</td>
</tr>
<tr>
<td><strong>Gross Domestic Expenditure</strong></td>
<td>1,368,069</td>
</tr>
</tbody>
</table>

Source: Statistics Canada, CANSIM table 380-0017.
Exercise 1.2: Can you think of an example of an individual who might belong to (i.e., makes purchases that would be reflected in) each of the four sectors described in Table 1.1?

The Income Approach

The income approach to computing the GDP relies on the following fact: *Every purchase of a good or service must constitute income to the agent (or agency) selling it.* The GDP calculated in this manner is sometimes referred to as the **Gross Domestic Income** (GDI). In principle, this can be done in any one of a number of ways.

The most obvious way to calculate aggregate income would be to define an individual (or household) as the basic economic unit. In general, each individual has multiple sources of income, including income on domestically-employed human capital $w$ (wages), income on domestically-employed capital $d$ (dividends, retained earnings, interest income on bonds), and income on assets employed in the foreign sector, $f$ (this could include wages earned outside of the country). In the computation of GDP, this latter source of income is left out (although, it is included in the measure of GNP). Thus, if there are $N$ domestic residents, one could compute:

$$GDI_t \equiv \sum_{i=1}^{N} (w_{it} + d_{it}).$$

Note that taxes and transfers to and from the government are not included in this measure. Why is this? Let’s think about it. Suppose that the government collects taxes $\tau_{it}$ from individual $i$. Note that this tax measure includes taxes from all sources, including sales taxes, property taxes, and income taxes (including corporate income, at least, on that fraction of the domestic business sector owned by domestic residents). What does the government do with these taxes? It uses them to pay the wages and salaries of government sector employees, which shows up in $w_{jt}$ for some government worker $j$. It also uses these taxes to make transfers to individuals in the economy. Note that these transfers do not count toward the GDP as they do not constitute any production of new good or service (they serve simply to redistribute output across members of society).

When one computes the GDP in this manner, it is interesting to note that most of the income generated in the economy accrues to human capital. For many economies, the income share of human capital ranges between 65–75%.

Naturally, this common-sense way of reporting income is not the way the IEA typically does things. Instead, the IEA reports the breakdown in income according to Table 1.2:
Table 1.2 Gross Domestic Product: Income-Based
Canada 2005 (millions of dollars)

<table>
<thead>
<tr>
<th>Category</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wages and Salaries</td>
<td>678,925</td>
</tr>
<tr>
<td>Corporate Profits (before tax)</td>
<td>193,936</td>
</tr>
<tr>
<td>Government Business Enterprises (before tax)</td>
<td>13,370</td>
</tr>
<tr>
<td>Net Interest Income</td>
<td>61,240</td>
</tr>
<tr>
<td>Net Income (farm)</td>
<td>1,551</td>
</tr>
<tr>
<td>Net Income (unincorporated farm)</td>
<td>84,666</td>
</tr>
<tr>
<td>Inventory Valuation Adjustment</td>
<td>-442</td>
</tr>
<tr>
<td>Taxes less subsidies on factors of production</td>
<td>59,961</td>
</tr>
<tr>
<td><strong>NDP (at factor cost)</strong></td>
<td>1,093,207</td>
</tr>
<tr>
<td>Taxes less subsidies on products</td>
<td>94,750</td>
</tr>
<tr>
<td>Capital Cost Allowance</td>
<td>181,427</td>
</tr>
<tr>
<td><strong>Gross Domestic Income</strong></td>
<td>1,369,384</td>
</tr>
</tbody>
</table>

Source: Statistics Canada, CANSIM table 380-0001.

As you can see from Table 1.2, it is somewhat cryptic (not to mention, amusing). Let’s try to figure out what’s going on here.

First, I am guessing that wages and salaries refers to wage income net of taxes. The reason I believe this is because there is also a category called taxes less subsidies on factors of production. Presumably, factors of production here refers only to human capital, as I notice that the income generated by the physical capital owned by the business and government sectors is reported on an after-tax basis. Note that corporate profits include earnings that are both retained and distributed as dividends, and is net of depreciation costs (which is why the capital cost allowance is added later on).

There is an issue here as to how the income of unincorporated (non-farm) businesses is treated. Presumably, this is just lumped in with wages and salaries, although properly speaking, at least a part of this should actually be recorded as capital income. It is rather amusing to see that there are separate categories for the farm sector (why?). Presumably, the net income from farms represents capital income. However, the net income of unincorporated farms includes both wage income and capital income (no attempt is made to separate out these components).

The NDP represents the net (of depreciation) domestic production. At factor cost means that the incomes were calculated net of ‘indirect’ taxes (like sales taxes). Thus, to compute the GDP, both capital consumption and indirect taxes must be added to the NDP figure.

The Income-Expenditure Identity

So far, we have established that \( \text{GDP} \equiv \text{GDI} \) and \( \text{GDP} \equiv \text{GDE} \). From these two equivalence relations, it follows that \( \text{GDE} \equiv \text{GDI} \). In other words, ag-
aggregate expenditure is equivalent to aggregate income, each of which are equivalent to the value of aggregate production. Again, the 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 \( GDI = GDE \) 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:

\[
Y = C + I + G + X - M. \tag{5}
\]

Note that since the income-expenditure identity is an identity, it always holds true. 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 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.

5. Other Measurement Issues

We’ve already talked a bit about some of the measurement problems concerning the classification of output into its consumption and investment components. That discussion, however, was predicated on the assumption that the concept of ‘output’ was well-defined and consistently measured. As we dig deeper into
the sausage-making machinery, however, we find that both of these assumptions need to be viewed with caution.

From the definition of GDP, we know—in principle, at least—that the GDP is supposed to represent some measure of the ‘value’ of what an economy produces in the way of (final) goods and services; or ‘output,’ for short. The easiest and most consistent way of aggregating the value of different goods (or factors of production) is by adding up their values on the basis of market prices—which are usually denominated in units of the national currency. Of course, such an exercise first presumes the existence of markets for different forms of output and factors of production; and second, presumes that prices and the quantities exchanged in these markets are somehow observable to national statistical agencies.

One might suppose that in the so-called developed world, that there a sufficient number of markets to price most forms of goods and services. But even in this best-case scenario, statisticians are confronted with a number of conceptual and practical problems.

The Government

Consider, for example, government ‘purchases’ of output. Some of this output is purchased from the private sector (e.g., military purchases from private defense contractors). But a large component of government ‘purchases’ is in the form of output that it produces itself and then transfers to the private sector at zero (or close to zero) prices; e.g., medical services, law enforcement services, education services, etc. How does one measure the market value of services such as these that are not sold on markets?

Since market prices do exist for private medical care and private education, one approach would involve trying to impute the value of this government production using the market prices of close substitutes available in the private sector. This method of imputation may not available for a large class of non-marketed goods and services, however (e.g., what is the market value of the services provided by the court system)? In practice, the way this is handled by the IEA is simply to assume that the market value of government production is equal to its factor cost. For example, if a judge is paid $150,000 per annum by the government, then this $150,000 figure is assumed to be the market value of the judicial services produced by the judge’s labor input.

This method of imputing the market value of government production may not be a bad approximation at most times and for most countries. Nevertheless, one should be aware of the potential pitfalls with this method. To see what can go wrong, consider the following stark example. Imagine an economy in which labor is the only factor of production and assume that workers are either employed in the private or government sector. The private sector produces and sells (at market prices) output $Y_p$ at the wage cost $W_p$, which yields profit
\[ \Pi_p = Y_p - W_p. \] The government produces output \( Y_g \) at the wage cost \( W_g \) and simply 'gives' \( Y_g \) away, financing its wage bill with a tax \( T \).

Using conventional IEA methods, the imputed value of government production is calculated as \( Y_g = W_g \). Calculating GDP by the expenditure approach in this economy would yield \( GDE = Y_p + Y_g \). Calculating GDP by the income approach would yield \( GDI = W_p + \Pi_p + W_g \). You should be able to convince yourself that \( GDE = GDI \).

Now, imagine in fact that the market value of \( Y_g \) is zero; that is, government production is considered to be a complete waste. In this case, the true GDP in this economy is given by \( Y_p \). However, the measured GDP continues to be \( Y_p + Y_g \). In this case, government production should in fact be counted as a transfer of resources (from private sector employees to government sector employees); it should not be counted as adding value to the economy as a whole.

Note that this type of problem is absent in computing the GDP generated by the private sector. Suppose, for example, that the business sector makes a 'mistake' in producing output that nobody values. In this case, the market value of \( Y_p \) is equal to zero. The expenditure-based GDP number is now given by \( Y_g \); and the income-based GDP number is now given by \( W_p + \Pi_p + W_g \). Since \( Y_p = 0 \), the business sector now makes a loss (negative profits) equal to \( \Pi_p = -W_p < 0 \). Hence, \( GDI = W_g \) and \( GDE = Y_g \), with \( GDI = GDE \) measuring the true value-added in this economy.

**Home Production**

Time-use studies reveal that out of the total amount of time available to individuals, only a relatively small fraction of this is devoted toward activities that produce goods and services that are sold in markets (or given away by governments). In other words, there is likely a significant amount of time (labor) devoted toward activities that generate home production, defined to be goods and services that are produced and consumed (or invested) within a household (and hence, not exchanged on any market). In addition to labor, households also generally have available a stock of capital that is likewise employed in home production (e.g., owner-occupied housing and consumer durables).

To see what sort of issues arise here, consider the following examples. Imagine that you own a house that you rent out at $1500 per month. Then this $1500 is counted toward the GDP, since it constitutes capital income for you (and generates $1500 per month worth of shelter services, which is consumed by your tenants). Imagine now that you decide to move into this home. Then you no longer report $1500 in rental income on your tax return. However, it still remains the case that the home is generating $1500 worth of shelter services that are now consumed by yourself. Consistency demands that the IEA impute the market value of these shelter services as valued-added. In fact, this is what many statistical agencies do.
In practice, however, statistical agencies often treat the valuation of home production inconsistently. Consider, for example, what happens if you own and operate an automobile. Your purchase of the automobile is recorded as an expenditure on a consumer durable. This durable good generates a flow of transportation services, the value of which should be recorded as valued-added. In fact, this is typically not the case. In contrast, if a taxi company purchases the same automobile, it is treated as investment and the revenue the taxi company earns from this asset is recorded as capital income.

Similarly, consider two households, one of which is ‘traditional,’ and the other which is ‘modern.’ By a traditional household, I mean one in which the father goes to work and spends a considerable amount of time around the home engaged in activities like mowing the lawn, cleaning the gutters, painting the house, repairing his automobile, etc. As well, the mother stays at home, raising the kids, cooking meals, cleaning the house, etc. By a modern household, I mean one in which both parents are employed and contract out extensively for services that ‘self-produced’ by a traditional family (e.g., they hire a nanny to cook, clean, and look after the kids; they hire contractors to effect home renovations, car repairs, and maintain the home, etc.).

Imagine that the two households described above are more or less similar in age structure, number of kids, education level of the parents, and other attributes. Then the true valued-added generated by each household is likely to be similar. But the measured value-added of the modern household is likely to be much higher, as many of its time-use activities are formally exchanged on markets (together with the fact that the IEA does not impute a value to household production in the form of raising kids, home maintenance, etc.).

The simple example described above warns us to be careful in interpreting time-series evidence of the growth in GDP as reflecting an increased level of production. In particular, to the extent that ‘household structure’ changes over time (from traditional to modern), much of the growth in GDP may simply reflect a measurement phenomenon (rather than reflecting true growth in production). A similar caveat is in order when one is making cross-country comparisons of GDP, especially between developed and underdeveloped countries. Much of the output that is produced in underdeveloped countries is likely to take the form of home production and hence not counted toward official GDP measures.

Exercise 1.3: According to this website: [www.globalissues.org](http://www.globalissues.org/ TradeRelated/ Facts.asp), half the world (nearly 3 billion people) manage to live on less than $2 a day. While these unfortunate souls are undoubtedly poor by any measure, explain why the $2 a day figure likely overstates the true extent of their poverty.

One might argue that conceptually, the GDP should measure the market value of ‘marketable’ output, even if it is not actually ‘marketed.’ That is, if I decide to clean the gutters of my home one fall morning, the IEA should (in
principle) count this cleaning service toward the GDP. While I did not market out for this service, it is clearly a service that I could have contracted out for.

This then raises the question of whether there are goods or services that are ‘non-marketable.’ In fact, one could argue that there are. An immediate example that comes to mind is sleep. One can produce and consume sleep and individuals clearly value sleep (a minimum amount of which is necessary to maintain the health of one’s human capital). It is impossible, however, to contract out for sleep services; i.e., I cannot get someone to sleep for me. The same might be said of learning. You cannot get someone to learn the contents of this text for you. Most forms of leisure activities appear to be non-marketable as well; e.g., having someone take a vacation on my behalf just doesn’t seem right.

Of course, one could—in principle, at least—attempt to impute a market value for non-marketable production as well. For example, the value of the time one spends producing leisure could be valued at the opportunity cost of this time (i.e., the wage that is foregone by consuming time in the form of leisure). For better or worse, the production/consumption of leisure is not viewed as contributing to GDP (even conceptually). Since people obviously do value leisure, however, the GDP cannot be considered the sole determinant of what determines individual well-being.

The Underground Economy

The underground economy refers to economic activity that is beyond the scope of government regulation and measurement. Underground activity typically takes place in well-defined markets, so that it is relatively easy to measure the market prices goods and services transacted in these markets. It is more difficult—if not impossible—however, to measure the volume of transactions, since they are purposely hidden. Underground activities may either be legal or illegal. If they are legal, they are hidden primarily for the purpose of evading taxes. For example, if you would like drywall installed in your basement, a drywall contractor may offer you two prices depending on whether you are willing to pay by cash or cheque (you will get a cheaper price if you pay with cash). If they are illegal, they are obviously hidden to avoid legal ramifications. In some jurisdictions, for example, purchasing and selling sex and certain forms of drugs is illegal.

By their very nature, underground economies are difficult to measure, so that any estimate of their size is necessarily imprecise. Nevertheless, some estimates do exist. According to one Economist Magazine article, for example, the underground economy in Italy is estimated to be between 15-27% of (measured) GDP. According to this article, underground businesses are widespread in agriculture, construction and services.

5www.economist.com/countries/Italy/profile.cfm?folder=Profile-Economic%20Structure
Once again, the presence of such unmeasured output should lead us to view official GDP numbers with a fair amount of caution as they likely understate the true value of production by a considerable margin. On the other hand, while the level of GDP may be understated, its growth rate may not be—at least, to the extent that underground (and other unmeasured) activity remains a relatively constant proportion of measured activity. Likewise, cross-country comparisons of GDP are likely to be more meaningful if the set of measured activities is more or less the same. We have no \emph{a priori} reason, however, to believe that either of these conditions are met in reality.

\begin{exercise}[1.4]
Imagine that a government suddenly enacts into legislation an oppressive tax regime on its citizens (e.g., a 100\% tax on all income). Explain why the economy’s measured GDP is likely to fall by much more than the true level of GDP.
\end{exercise}

6. 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 \emph{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 \textbf{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 above). 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 (with some notable exceptions), the general level of prices tends to grow over time; such a phenomenon is known as \textbf{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}\times$100$ = $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 \((1.10)^{10}\times(3635) = 9451\) 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 describe one approach that is commonly adopted by statistical agencies. Following the discussion surrounding the expenditure-based GDP measure, we have GDP given by (1); which I reproduce here for convenience:

\[
GDE_t \equiv \sum_{i=1}^{Q} p_t^i x_t^i.
\]

As this measure relies on current (i.e., date \(t\)) prices (whether actual or imputed), it is sometimes referred to as the GDP measured at current prices; or
simply, the nominal GDP.

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

\[
RGDP_t \equiv \sum_{i=1}^{Q} \frac{Q_i^{1997}}{P_i^{1997}} x_i^t.
\]

This measure is called the GDP measured at base year 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, \( RGDP_{1997} \equiv GDE_{1997} \).

Figure 1.2
Nominal vs. Real GDP
Canada 1961-2005

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 ‘average’ level of prices 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.6 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.

Real per capita GDP

In general, the real GDP of any economy may rise (or fall) owing to: [1] a

---

6 Some statistical agencies have introduced various ‘chain-weighting’ procedures to mitigate this problem.
rise (or fall) in population; and/or \[2\] a rise (or fall) in the output produced per person. To get a sense of how material living standards for the ‘average’ person in an economy, it makes sense to divide an economy’s total real GDP by population size. The resulting number is called the real per capita GDP; or more commonly, the real per capita income.

It cannot be stressed enough that extreme caution should be exercised in interpreting real per capita GDP as a measure of material living standards or as a measure of economic welfare. First, one should keep in mind all of the measurement issues discussed at length above. Second, as far as material living standards are concerned, theory suggests that consumption is likely to constitute a better measure (or perhaps even wealth, to the extent that consumption is related to wealth). One would not want to judge the material living standards of a student with zero income, for example, solely on the basis of his or her income. Finally, there is good reason to believe that economic welfare depends not only on one’s consumption flow, but also on other things (e.g., leisure time spent with one’s family and friends).\(^7\)

With the appropriate caveats in place, let us examine the behavior of real per capita GDP for Canada. Figure 1.4 plots the evolution of Canada’s population and Figure 1.5 plots the real per capita GDP.

Figure 1.4
Population of Canada
1961:1 - 2005:1

Figure 1.5
Real per capita GDP
Canada 1961:1 - 2005:1

Source: Statistics Canada, table 510005
Source: CANSIM II, Series V1992067 (divided by population).
7. Growth and Business Cycles

The pattern of economic development displayed in Figure 1.5 for Canada is typical among many countries, especially for those that occupy the so-called ‘developed world.’ The most striking feature of Figure 1.5 is the trend rate of growth in real per capita income. In 1961, income per capita was approximately $13,100. By the end of the sample in 2005, income per capita had grown by a factor of 2.7 (to $35,600). This represents an average annual growth of approximately 2.2%.

Now, 2.2% may not sound like a large number to you. And you may be tempted into thinking that it really does not matter very much whether an economy grows at 1.2%, 2.2%, or 3.2%. In fact, even seemingly small differences in long-run growth rates such as these can translate into huge differences in the level of income over time. The reason for this lies in the power of compound interest.

To appreciate the power of compound interest, imagine that there are three economies A, B, and C that are currently generating $10,000 in per capita income. Economy A grows at \( g = 1.2\% \), economy B at \( g = 2.2\% \), and economy C at \( g = 3.2\% \) per annum. What will be the level of per capita income at the end of 20 years? The answer is provided in Table 1.3.

<table>
<thead>
<tr>
<th>Initial GDP</th>
<th>GDP after 20 years</th>
<th>Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Economy A ((g = 1.2))</td>
<td>$10,000</td>
<td>$12,694</td>
</tr>
<tr>
<td>Economy B ((g = 2.2))</td>
<td>$10,000</td>
<td>$15,453</td>
</tr>
<tr>
<td>Economy C ((g = 3.2))</td>
<td>$10,000</td>
<td>$18,776</td>
</tr>
</tbody>
</table>

A useful formula to keep in mind is the so-called Rule of 72. This rule states that if an economy grows at a rate of \( g\% \) per annum, then the number of years it takes to double income is given by \( n \):

\[
 n = \frac{72}{g}. \tag{6}
\]

Thus, an economy that grows at 2% per annum will increase its living standards by 100% in 36 years. An economy that grows at 4% per annum will double its living standards in only 18 years.
Any factor (including government policy) that may affect an economy’s long-run growth rate by even a small amount can ultimately result in very large differences in living standards over prolonged periods of time.

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.

Traditionally, macroeconomics has concerned itself more with the issue of ‘short run’ growth, or what is usually termed 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 1.5, we see that while per capita GDP tends to rise over long periods of time (at least, in Canada and some other countries), 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 (i.e., the shaded regions in Figure 1.5).

It is important to keep in mind that while it is tempting to dichotomize a pattern of economic development like Figure 1.5 into its ‘trend’ and ‘cycle’ components, there is in fact no a priori reason to believe that such a decomposition makes any sense theoretically (although, one could certainly perform such a decomposition statistically). Because what I’ve said here is important and not widely appreciated, let me elaborate.

When viewing a diagram like Figure 1.5, the natural inclination is to draw a smooth line through the data and interpret this as ‘trend.’ The difference between the actual data and the trend line is then interpreted as ‘cycle.’ Unfortunately, there is no unique or obvious way to detrend time-series data. In Figures 1.6, I display the trend and cycle components of Canadian GDP using one particular method (a cubic trend).
There is nothing inherently wrong in detrending time-series data. The mistake that people commonly make, however, is to assume a smooth trend line—estimated on historical data—actually represents a trend that can be expected to prevail into the foreseeable future. To put things another way, people commonly make the mistake of assuming that a smooth trend line represents something ‘real’ or ‘fundamental’ about the way an economy functions. In fact, a smooth trend line may be nothing more than a statistical illusion.

To make what I am saying more concrete, consider the following argument. Let \( y_t \) denote the (log) real per capita GDP and assume that you and I know underlying data generating process (DGP) for the economy. Imagine further that this DGP is given by:

\[
y_{t+1} = \gamma + y_t + e_{t+1},
\]

where \( e_{t+1} \) is a random variable, representing an unforecastable ‘shock’ to the economy’s GDP. We can assume that \( e_{t+1} \) takes one of two values, each of which is determined by the flip of a coin. Or, to be slightly more sophisticated, we can assume that \( e_{t+1} \) is determined by a draw from a Normal distribution with mean \( \mu \) and standard deviation \( \sigma \). If we set \( \mu = 0 \), then the expected value of \( e_{t+1} \) as of date \( t \) is given by \( E_t e_{t+1} = 0 \). In fact, the expected value of \( e_{t+j} \) for any \( j = 1, 2, 3, \ldots \) is given by \( E_t e_{t+j} = 0 \).

The DGP in (7) is an example of what is called a random walk with drift. The ‘drift’ parameter \( \gamma \) represents the expected rate of growth of (log) GDP. We
can simulate an equation like (7) by assigning parameter values and generating the ‘shocks’ \( e_t \) using a random number generator. Suppose, for example, that we set \( y_{1961:1} = 9.48 \) (its value for the Canadian economy in 1961). Let \( \gamma = 0.0055 \) (which generates an annual expected growth rate of 2.2\%). As well, choose a standard deviation \( \sigma = 0.015 \). Figure 1.7 displays two simulated time-series for \( y_t \) (along with the actual Canadian data).

One thing that should strike you from viewing Figure 1.7 is how the simulated series resemble the actual data; in fact, it would be very hard to tell (without knowing beforehand) which series was generated by our model (7) and which was generated by the economy.

![Figure 1.7](image)

Log GDP: Actual and Simulated

Now, imagine that we handed our simulated data to an econometrician and asked him or her to estimate a smooth trend line the way that was done in Figure 1.6. One could certainly do this; and we would be left with a diagram similar to Figure 1.6. But does the estimated trend line represent anything ‘fundamental’ about the manner in which our model economy functions? In particular, is there any reason to believe that future GDP levels will revert back to the estimated trend line? Can we use the estimated trend line to forecast the future level of GDP? The answers to these questions is no.

The reasons for why we can be so certain of this are twofold: [1] unlike the econometrician, we know the true DGP generating our simulated series; and [2]
given that we know the true DGP, we can compute the theoretical trend and show that does not correspond to a smooth line.

The theoretical (i.e., true) trend behavior displayed by our model economy can be calculated as follows. Using (7), we can deduce that:

\[ y_{t+2} = \gamma + y_{t+1} + e_{t+2}; \]
\[ = 2\gamma + y_t + e_{t+1} + e_{t+2}. \]

Similarly,

\[ y_{t+3} = \gamma + y_{t+2} + e_{t+3}; \]
\[ = 3\gamma + y_t + e_{t+1} + e_{t+2} + e_{t+3}. \]

Continuing on in this way, we have:

\[ y_{t+n} = n\gamma + y_t + e_{t+1} + e_{t+2} + ... + e_{t+n}, \]

for any arbitrary \( n > 0 \).

Now, suppose we are at date \( t \) and wish to estimate the future level of GDP at date \( t + N \), where \( N \) is some number large enough to be considered the ‘long-run;’ e.g., \( N = 40 \) quarters (ten years) off into the future. Then we can ask the question: what is the expected value of \( y_{t+N} \), given what we know at date \( t \)? The answer is given by:

\[ E_t y_{t+N} = N\gamma + y_t. \] (8)

This ‘long-run’ (\( N \)-period ahead) forecast of GDP can be thought of the model’s ‘trend’ for the (log) level of GDP. In other words, the ‘long-run’ level of GDP is determined by the current level of GDP \( y_t \) plus an expected growth term \( N\gamma \).

Note that since \( y_t \) changes over time, so does the trend level of GDP. In fact, since \( y_t \) is a random variable, our model displays what is known as a stochastic trend. Needless to say, a stochastic (random) trend line is not going to look anything like the smooth deterministic trend line drawn in Figure 1.5. There is, in fact, no reason to believe that \( y_t \) will ever revert back to a smooth trend line estimated with historical data. The estimated trend line is simply a figment of the econometrician’s imagination (we know this, but he doesn’t).

Why is this important to understand? It is important for the following reasons. First, as mentioned earlier, there is a tendency for people to believe that a smoothly drawn line through the data represents something ‘fundamental’ about the way an economy functions in the long-run. In other words, there is a tendency to believe that an economy will eventually revert back to some given trend behavior. As the example above demonstrates, such a belief is not necessarily correct (and can easily be incorrect, given how similar the simulated series in Figure 1.7 resemble the actual data).

Second, by assuming that the economy does possess a smooth trend, one is implicitly assuming that growth and business cycle phenomena are independent
of each other. That is, one is easily led to the conclusion that we can use a growth theory to understand trend behavior and a business cycle theory to understand deviations from trend, with each theory bearing no relation to one another. In fact, it may very well be the case that the so-called ‘business cycle’ is nothing more than a by-product of the process of economic development (as suggested by our model). In other words, we may be wrong in thinking (as people commonly do) that we can understand the business cycle without understanding the process of growth itself.

Fluctuations in GDP may largely be a by-product of a random growth process (a shifting trend); i.e., the ‘business cycle’ may be inextricably linked to the process of economic development itself.

The basic lesson here is to be careful in assuming that an economy has a smooth trend and that GDP will eventually return to this trend. To demonstrate the potential pitfall of this commonly held view, consider Figure 1.8, which plots the real per capita GDP for Japan from 1960–2004. Imagine that an econometrician living in the year 1973 wants to estimate the ‘trend’ for Japanese GDP based on the historical data 1960-73. The dashed line (a simple linear trend) in Figure 1.8 appears to fit the historical data reasonably well. Unfortunately, the forecast of GDP far off into the future would have been off a tad.
8. 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. It should come as no surprise then to learn that there are many different hypotheses that offer different interpretations of observed patterns. At the end of the day, the lines of debate are drawn across the following two questions:

- What are the primary shocks that are the ultimate source of aggregate fluctuations?
- What is the mechanism by which an economy responds to any given shock?
Strictly speaking, a shock refers to a ‘surprise’ event that is determined by God or nature (i.e., an event that is beyond the control of any economic agent or agencies). A tsunami that wipes out a significant fraction of a region’s stock of human and physical capital constitutes a possible example. The sudden appearance of new technology—like the internet—may be another.

Unfortunately, the interpretation of shock events is not entirely unambiguous. For example, some religious groups contend that the December 26, 2004 tsunami that afflicted southeast Asia was in fact brought forth by God/Allah/Jehovah as a punishment for the region’s sins (sex and drugs). According to this interpretation, the inhabitants (and tourists) in southeast Asia brought the tsunami on by their own debaucherous behavior. This view requires that we take as exogenous (i.e., unexplained) two things: [1] God’s law; and [2] preferences for debaucherous activities (that violate God’s law). Nevertheless, one might still argue that while the tsunami itself should have been expected, the exact date of its arrival could not have been forecasted. Thus, the actual arrival of the tsunami is still usefully interpreted as a shock.

The same sort of argument can be made with respect to a technology shock. That is, let us take as exogenous two things: [1] the law of nature governing the process of discovery; and [2] preferences for higher living standards. Then one might reasonably argue that the idea behind the internet was in fact the product of human behavior (e.g., R&D activity). As with the tsunami, however, no one can reasonably be expected to forecast the exact arrival date of any technological advancement. When knowledge is discovered then, it comes as a shock.

In a sense, any economic theory constitutes an explanation of how a set of endogenous variables $Y$ is determined in relation to a set of exogenous variables $X$. Thus, in abstract terms, any theory can be thought of taking the following form:

$$X \rightarrow_{L} Y,$$

where $\rightarrow_{L}$ denotes the logic underlying the explanation. A shock then can be thought of as some exogenous change in $X$ and denoted $\Delta X$. The theory then provides an explanation for the mechanism by which a change in $X$ might be expected to influence the endogenous variables $Y$; i.e.,

$$\Delta X \rightarrow_{L} \Delta Y.$$  

At issue then is what to include in $X$ and how to think about $\Delta X$ (assuming that few people will argue with $\rightarrow_{L}$ or the form of logic to be used in connecting assumptions with predictions). Ultimately, one would hope for a theory that could explain everything in terms of an $X$ that was ‘truly’ exogenous. Unfortunately, the nature of economics (and of science in general) is such that a ‘grand unifying theory’ of this form is unlikely to found anytime soon. In the meantime, we have to make due with what must be considered only ‘partial’ explanations that will (hopefully) be improved upon as the science progresses.
Thus, there is at present a long list of candidates for what might be included in $\Delta X$. Some theories assert the existence of government spending shocks or monetary policy shocks, as if the behavior of the government or its affiliated agencies is beyond comprehension (i.e., determined by God or nature). The Bank of Canada, for example, emphasizes domestic shocks that arise from political uncertainty (e.g., will Québec separate from Canada or not) and international shocks like the 1973 OPEC oil crisis and the 1997 Asian financial crisis. Economic commentators and analysts on television are fond of pointing to price shocks (e.g., the stock market, interest rates, inflation, exchange rates) as if these objects too are somehow not determined by conscious human behavior in reaction to more fundamental disturbances.

In many cases, it can make sense to view particular events such as sudden price changes or financial crisis as an exogenous shock, even if we know (or suspect) that prices and financial market behavior are not truly exogenous. For example, we may want to frame a question in the following way: How might the domestic economy react given the financial crisis in Asia? The answer to such a question, while useful for some purposes, is ultimately unsatisfying as it leaves unexplained the crisis itself. Economists have different hypotheses concerning the ultimate source of such disturbances, and these different views help define various schools of thought.

At the risk of oversimplifying, one might usefully categorize macroeconomic theory into two broad schools of thought, each of which is characterized primarily by the particular set of shocks and mechanisms that tend to be emphasized. I label the first school conventional wisdom, as variants of this view are held so widely among market analysts, politicians, central bankers, and a good part of the academic community. I label the second school neoclassical; this view is not nearly so widely-held, but is nevertheless influential among academic economists.

Conventional Wisdom

The conventional wisdom owes its intellectual debt primarily to the work of John Maynard Keynes, whose views on the business cycle were shaped to a large extent by the events of the Great Depression. The primary legacy of this view is twofold: [1] that shocks are ultimately the result of exogenous changes in private sector expectations (animal spirits); and [2] that market economies are sufficiently dysfunctional as to make well-designed government stabilization policies desirable.

The way these ideas have evolved into conventional wisdom are as follows. First, growth is explained as the product of a relatively smooth process of technological development, so that one can infer from the data a relatively stable
trend that determines an economy’s ‘long-run’ fundamentals. The business cycle, then, constitutes fluctuations around this trend (with movements in GDP eventually reverting back to trend). The trend level of GDP is sometimes referred to as supply or potential GDP, with the actual level of GDP referred to as demand.\textsuperscript{10}

Having identified a relatively stable trend (supply) and then observing that actual GDP (demand) fluctuates around trend, one is led to the conclusion that business cycles are caused by demand shocks (i.e., unexplained and random changes in desired spending patterns emanating from various sectors of the domestic and foreign economy). While the root cause of these shocks is not usually discussed, it seems clear enough from the language used to describe them that they are thought to be the product of exogenous (and irrational) swings in market sector expectations (animal spirits). A strong quarter, for example, might be explained as resulting from the ‘strength of the consumer,’ which in turn may lie in the behavior of ‘consumer confidence’ (high expectations of future earnings). Similarly, business sector behavior may be described as being the product of ‘irrational exuberance’ (high expectations concerning the future return to investment).

To the extent that demand shocks are ‘irrational,’ they have adverse consequences that can last a long time. A bad investment today, for example, will have implications for GDP many periods into the future. These shocks are further exacerbated by various market imperfections—primarily in the form of ‘sticky’ nominal prices and wages—that prevent markets from adjusting rapidly to shocks (which explains why ‘supply’ is not usually equated to ‘demand’). 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.\textsuperscript{11}

**Neoclassical View**

The neoclassical view is closer in spirit to those expressed by another great economist, Joseph Schumpeter.\textsuperscript{12} The primary legacy of this view is that technology shocks—the very shocks that contribute to the general rise in living standards—are at the same time responsible for generating the fluctuations that are commonly interpreted as the business cycle.

\textsuperscript{10}This language has even found its way into the IEA. For example, the expenditure-based measure of GDP is often labelled ‘final demand.’ The implication, of course, is that the final demand computed in this manner does not necessarily measure ‘final supply.’

\textsuperscript{11}It is interesting to note that Keyne’s (1936) own views differ significantly from those that evolved from his work. In particular, while he emphasized the role of ‘animal spirits,’ he viewed these exogenous changes in expectations as being rational in the sense of constituting ‘self-fulfilling prophesies.’ Further, the concept of ‘sticky’ prices or wages played no role in his theory; except peripherally and as a mechanism that potentially dampened the adverse consequences of demand shocks.

According to the neoclassical view then, 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 conventional wisdom, which views trend growth as being relatively stable, the neoclassical view asserts 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 may 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). Further, there is no guarantee that all new technologies work out exactly as planned or expected. What may have looked promising at one date, may in fact turn out to be a disaster later on (resulting in an observed negative technology shock).

As with the conventional wisdom, the neoclassical view admits that sudden changes in private sector expectations may lead to sudden changes in desired household 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, or other real factors). In other words, changes in market sentiment are the result and not the cause of the business cycle. It is important to keep in mind when evaluating this perspective that the concept of ‘rational’ expectations does not imply that individuals never make ‘mistakes.’ It simply means that expectations are formed in the ‘best’ way possible, using whatever relevant information is currently at one’s disposal. More often than not, actual outcomes will differ from those that are expected.

According to the neoclassical view, the business cycle is an unfortunate but largely unavoidable product of the process of economic development. Market imperfections play little or no role in exacerbating economic fluctuations; indeed, even a well-functioning ‘planned’ economy (if such an object were to exist) would exhibit similar fluctuations. Given this interpretation, 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.
9. The Plan Ahead

In the chapters that follow, I plan to lay out—in easily digestible bits and pieces—the various key elements that constitute modern macroeconomic theory primarily from the neoclassical perspective. This endeavor is not meant to be an exercise in pure theory; throughout the book I try to demonstrate how the theory can be used to interpret and understand various aspects of real-world economies.

I do spend some time explaining aspects of what I have called the conventional wisdom. I do this primarily because it is impossible to understand the rationale behind many current and past government policies without reference to this view. I do not, however, spend very much time developing the theoretical underpinnings of the conventional view, much of which is embodied in the so-called New-Keynesian school of thought. The reasons for proceeding the way I do are as follows:

- There are already many textbooks out there that exposit the conventional wisdom (albeit, at a highly rudimentary level); there is no reason for me to add yet another text regurgitating the same basic ideas and methods in a slightly different way;
- By the same token, there are relatively few modern treatments of the neoclassical perspective at the undergraduate level—which I view as glaring gap, given the growing influence this approach is having on the profession;
- I believe that a thorough understanding the neoclassical perspective constitutes a prerequisite for a deep understanding of the conventional view. Keep in mind that Keynes’ own General Theory was meant to be a generalization of classical theory; likewise, at the core of New-Keynesian theory lies a neoclassical model extended or modified along various dimensions.

The book is designed so that, by the end of it, a conscientious (and patient) reader will have a reasonably good idea of basic neoclassical theory (and how the various bits and pieces I present can ultimately be tied together in a more advanced theoretical treatment), together with an idea as to how neoclassical theory can be applied toward interpretation, prediction, and the evaluation of policy.
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. Explain why ‘overpaid’ government employees will lead to an overstatement of GDP, whereas ‘overpaid’ private sector employees will not.

3. 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 $B$?

4. Consider the following commentary by Michael Mandel, in an article entitled *Why the Economy is a Lot Stronger than you Think*. Based on what you have learned in this chapter, can you spot all the errors he makes?13

---

13See: http://www.businessweek.com/magazine/content/06_07/b3971001.htm
A central feature of the business cycle is the comovement between output (real per capita GDP) and employment (or hours allocated to work activities per capita). In the short-run, output and employment tend to move in the same direction. In fact, much of the change in GDP over the cycle is attributable to changes in the level of employment. This makes a lot of sense since, as more individuals work to produce output (or as employed individuals work longer hours), one would expect the level of output to increase. But understanding this fact alone does not help us understand the business cycle, as it does not explain why employment should change in the first place.

Most economists would probably agree that the cyclical variation in employment is driven by fluctuations in the demand for labor. There is, unfortunately, considerably less agreement on what forces are responsible for generating the volatility in labor demand. Since labor productivity and real wages tend to be procyclical (i.e., move in the same direction as GDP), some economists stress the role played by productivity shocks (recall the discussion in Chapter 1). The basic idea here is that temporal variation in the productivity of labor is a natural phenomenon in a growing economy. When productivity is high (relative to trend), the business sector demands more labor to exploit the high return to labor. This shift in labor demand puts upward pressure on real wages, which serves to draw more individuals into the labor force. The reverse holds true when productivity is low (relative to trend).

The primary goal of this chapter is to formalize the intuition above in terms of an explicit (i.e., mathematical) theory. Developing a formal model will prove useful for a number of reasons. First, it will allow us to check whether the intuition expressed above survives a logical analysis. (There are times when simple intuition only holds under some very specific conditions—or perhaps not at all). Second, we can use the logic contained in the theory to help us evaluate the potential role for government policy. Third, the simple theory developed here will serve as useful groundwork for the more elaborate theories to be developed later on in the text.

To this end, we will construct a model economy, populated by individuals that make economic 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. If these trade-offs fluctuate over time owing to any sort of exogenous shock, then individuals are likely to change their behavior accordingly. The question here is whether exogenous changes
in productivity might generate changes in behavior that imply business cycle activity that is qualitatively similar to what is observed in reality (i.e., procyclical employment and real wages).

2. A Simple Model

The model I consider here is highly simplified indeed—but one must learn how to crawl before learning how to walk and then run. Do not be put off by the starkness of the model—this is done primarily for pedagogical reasons. You can rest assured of two things: [1] the model can be easily extended in a number of interesting (and more complicated) ways—some of which we will explore later on; and [2] the basic forces highlighted in the simple model continue to hold in much more general (and realistic) environments.

We begin by stating a number of simplifying assumptions. Since employment behavior plays an important role in the business cycle, we want to think of how to model the labor market. To this end, we want to model a household sector (from which stems the supply of labor) and a business sector (from which stems the demand for labor). So to begin, let us assume that the economy consists only of these two sectors; i.e., assume that there is no government or foreign sector. From our knowledge of the income-expenditure identity, we know that this assumption implies \( G = X = M = 0 \), so that \( C + I = Y \). If we assume further that all output is in the form of consumer goods and services, then \( I = 0 \) and \( C \equiv Y \). In other words, all income in this model will take the form of claims against domestically-produced consumer goods and services. In short, we are dealing here with a closed economy, with no government, no foreign sector, and no investment. These assumptions will be relaxed in later chapters.

Let us think next of the people that occupy our hypothetical world. We want to think of an economy consisting of a ‘large’ number of people, each of whom belong to the household sector. In reality, people obviously differ along many dimensions. On the other hand, people also seem to share many things in common, including a general desire to advance their material well-being. Our strategy here will be to focus on these common attributes and downplay the differences. For simplicity, we take this to the extreme by assuming the existence of a representative household (i.e., we assume that households are all identical along economically relevant dimensions).\(^{14}\)

Let us now think about the business sector. We want to think of the business sector as consisting of a ‘large’ number of competitive firms. It is important to note that firms are not people; they are simply legal entities (operated by people).

---

\(^{14}\text{Again, the student should note that one could easily extend the model to incorporate heterogeneity among households. The primary cost of doing this is some added mathematical complexity (we would have to keep track of distribution functions); see, for example, Appendix *****.} \)
that organize production. Again for simplicity, we will assume that all firms are identical so that there exists a representative firm. Assume that firms are owned by members of the household sector (to which all individuals belong) and that firms are motivated by a desire to maximize shareholder wealth.

Finally, we want to consider an assumption that will simplify decision-making considerably. In particular, we consider here what is called a static model. 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 focussed on here have no intertemporal dimension (which allows us to abstract from financial markets). The restriction to static decision-making allows us, for the time-being, to focus on intratemporal decisions (such as the allocation of time across competing uses over the course of a year). As such, one can interpret the economy as generating a sequence of static outcomes over time.

2.2 The Household Sector

The representative household has preferences defined over two objects: [1] a basket of consumer goods and services (consumption), which we denote by c; and [2] a basket of home-produced goods and services (leisure), which we denote by l.¹⁵ Let (c,l) denote a commodity bundle (i.e., a particular quantity of consumption and leisure); this is also sometimes called an allocation. The set of all conceivable commodity bundles (allocations) is called the commodity space.

Household preferences are defined over the commodity space. What this means in plain language is that we assume that households can rank different commodity bundles by making statements like: I prefer (c₁,l₁) to (c₂,l₂), or I am indifferent between (c₁,l₁) and (c₂,l₂). Under some weak conditions, we (as theorists) can represent such preferences with a mathematical relation called a utility function. In particular, let u(c,l) denote the rank attached to any given commodity bundle (c,l). Then the statement I prefer (c₁,l₁) to (c₂,l₂) can be represented by u(c₁,l₁) > u(c₂,l₂) and the statement I am indifferent between (c₁,l₁) and (c₂,l₂) can be represented by u(c₁,l₁) = u(c₂,l₂). This is not rocket-science.

Note that by specifying the household’s preferences explicitly (by way of a utility function), we are being very explicit about what motivates household behavior. We are suggesting that households care about the level of broad-based living standards, both in the form of consumption and leisure. Given a choice, households will presumably choose the (c,l) that they rank most highly; i.e., households are motivated by the desire to maximize utility u(c,l). In plain language, we are just assuming that households desire to do the best they can according to their preferences. This is not an unreasonable assumption; and in

¹⁵Note that the value of home-produced output (leisure) is not counted as a part of the GDP.
To gain predictive power, we need to make a few (standard) assumptions regarding preferences. First, assume that more is preferred to less, so that the utility function is increasing in both $c$ and $l$. Second, assume that both $c$ and $l$ are normal goods. What this will imply is that as a household becomes wealthier, it will demand more of both $c$ and $l$ (holding the relative price of these two goods fixed). Third, assume that preferences are transitive. What this means is that a household prefers $A$ to $B$ and $B$ to $C$, then it also prefers $A$ to $C$. Finally, assume that the utility function is continuously differentiable and that it is strictly concave in each argument. This latter assumption implies that the household experiences a diminishing marginal utility of consumption and leisure as their levels are increased. Taken together, these assumptions allow us to represent preferences diagrammatically in the commodity space by way of indifference curves; i.e., see Figure 2.1.

In Figure 2.1, $u_0$ and $u_1$ simply represent two different numbers that assign a ranking to each point on their respective curves. If we fix a utility-ranking at some number $u_0$ (e.g., $u_0 = -3$), then the associated indifference curve is defined

---

16 One alternative, employed in evolutionary economics, is to assume that people are born with pre-programmed behavioral rules subject to the forces of natural selection.
to be all the combinations of \((c, l)\) that generate this rank; i.e., \(u_0 = u(c, l)\).

By construction, a household is indifferent between all the commodity bundles located on a given indifference curve. Note that the commodity bundles located on the indifference curve associated with \(u_1\) are preferred (i.e., yield higher welfare) to those located on the lower indifference curve; i.e., \(u_1 > u_0\).

Transitivity implies that the indifference curves of a given utility function can never cross. (Prove this).

A concept that we will make great use of throughout the text is the so-called **marginal rate of substitution** (or MRS, for short). Essentially, the MRS refers to the **slope of the indifference curve** at any point in the commodity space (actually, it is the absolute value of this slope). Notice that the MRS is a function, since it depends on where one is positioned on the indifference curve; to emphasize this, we can write \(MRS(c, l)\).

The MRS has an important economic interpretation. In particular, \(MRS(c, l)\) provides a measure of the household’s relative valuation of consumption and leisure at any bundle \((c, l)\). For example, consider some allocation \((c_0, l_0)\) with associated 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 \(\Delta l\) 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 \(\Delta c\) 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 \(\Delta c\) that satisfies the following condition:

\[
u_0 = u(c_0 + \Delta c, l_0 - \Delta l).
\]

For a very small \(\Delta l\), the number \(\Delta c/\Delta l\) 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 \(\Delta c/\Delta l\) 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 \(\Delta c/\Delta l\) is a small number.

Is your head hurting right about now? Just sit back and relax for a moment. What we have done so far is very simple. We’ve assumed a world populated a large number of identical households. Each household cares about its material living standard, in the form of consumption and leisure. Generally speaking, the more the better. But households are also willing to substitute consumption for leisure (and vice-versa). You might be willing to sacrifice a lot of leisure in exchange for living in a big house, for example. How willing you are to do this depends on the nature of your preferences. We (as theorists) can represent this willingness by way of a diagram with appropriately shaped indifference curves (with the shape influencing the MRS). Furthermore, we can use these
indifference curves to ascertain which allocations are preferred relative to others. If your head is still hurting, try a couple of extra strength Tylenol.

Before I stop talking about preferences, I want to stress once more why we (as theorists) go through all of the trouble of modeling them. 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. All we really need to know (at least, for making qualitative, rather than quantitative 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.

By modeling preferences explicitly, we take a stand on what ultimately motivates decision-making. Information concerning preferences can then be used to help predict behavior and evaluate the welfare consequences of policy and other exogenous events.

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:

\[ n + l = 1. \]  

(9)

Equation (9) is referred to as a time constraint. 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.
In the model, households will generally have two sources of income: labor income and non-labor income. Let \( w \) denote the real wage (i.e., the amount of output that can be purchased with one unit of labor), so that labor income is given by \( wn \). Denote non-labor income by \( d \) (i.e., the dividends that would be accruing to the household sector via their ownership of firms in the business sector). For now, we simply view \((w, d)\) as parameters (i.e., exogenous variables) that are beyond the control of the household. Later, I will describe how these variables are determined by market forces.

Since this is a static model, all income earned is consumed (i.e., none of it is saved). The household’s budget constraint is therefore given by:

\[
c = wn + d.
\]

By combining the time constraint (9) with the equation above, we can rewrite the budget constraint in terms of consumption and leisure:

\[
c + wl = w + d. \tag{10}
\]

In the equation above, \( w + d \) is sometimes referred to as full income; i.e., the combined value of the household’s time endowment and non-labor income. Out of this full income, the household makes purchases of consumption and leisure, with the price of leisure (measured in units of consumption) given by the real wage.

I remarked earlier that the shape of the indifference curve (in particular, the MRS) reflects the household’s willingness to substitute consumption for leisure. Observe now that the budget constraint reflects the household’s ability to substitute consumption for leisure. The interaction between the willingness and ability to substitute across commodities is a concept that plays a central role in economic analysis, so it will be helpful to keep it in mind always.

Now that we have described what motivates and constrains household choices, we are in a position to deduce their behavior. Consider an arbitrary \((w, d)\), which the household views as beyond its control. Given this \((w, d)\), the household is assumed to choose its most preferred allocation \((c, l)\) that at the same time respects its budget constraint. In mathematical terms, the choice problem can be stated as:

Choose \((c, l)\) to maximize \(u(c, l)\) subject to: \(c + wl = w + d\).

This problem has a solution (representing the household’s optimal choice). Without saying what the solution is, we can denote it by a pair of choices \((c^D, l^D)\), where \(c^D\) can be thought of as ‘desired consumer spending’ (consumer

\(^{17}\)Note that the income that workers earn here is in the form of privately-issued claims against output. Think of these claims as coupons that are redeemable in merchandise.
demand) and \( l^D \) represents the demand for leisure. Note that since the total time endowment is fixed, the demand for leisure automatically implies a supply of labor, \( n^S = 1 - l^D \). In terms of a diagram, the optimal choice is displayed in Figure 2.2 as allocation A (This figure is drawn for the case in which \( d = 0 \)).

**FIGURE 2.2**
Household Choice

![Diagram showing household choice with budget line, consumption (c), labor (l), and leisure (l^D) axes.](image)

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 \( w + d \) (with \( d = 0 \) here). 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 con-
ditions at the same time:

\[ MRS(c^D, l^D) = w; \]  
\[ c^D + wt^D = w + d. \]  

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.

Notice that since we have assumed that the household makes its choice conditional on some prevailing pattern of wages and dividends \((w, d)\), it follows that the optimal choice will, in general, depend on these parameters. At times, we may wish to emphasize this dependence by writing the solution explicitly as a function of the underlying parameters; e.g., \( c^D(w, d), l^D(w, d) \) and \( n^S(w, d) \).

**Exercise 2.2:** Suppose that preferences are given by the utility function \( u(c, l) = \ln(c) + \lambda \ln(l) \), where \( \lambda > 0 \) is a preference parameter. For these preferences, one can demonstrate that \( MRS(c, l) = \lambda c/l \). Use this information, together with the conditions in (11) to solve explicitly for consumer demand, the demand for leisure and (from the time constraint) the supply of labor.

The theory developed here makes clear that the allocation of time to market work (the supply of labor) should depend on the return to work relative other potential uses of time (in this simple model, leisure is the only alternative). The return to work here is given by the real wage. Intuitively, one would expect that an exogenous increase in the real wage might lead a household to reduce its demand for leisure (and hence, increase labor supply). While this intuition is not incorrect, it needs to be qualified. We can discover how by way of a simple diagram.

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 (again, drawn for the case in which \( d = 0 \)) shows that the household may respond in three general ways, represented by the allocations B, C, and D. In each of these cases, consumer demand is predicted to rise. However, the effect on labor supply is, in general, ambiguous. Why is this the case?

An increase in the real wage has two effects on the household budget. First, the price of leisure (relative to consumption) increases. Our intuition suggests that households will respond to this price change by substituting into
the cheaper commodity (i.e., from leisure into consumption, with the implied increase in labor supply). This is called the substitution effect. Second, household wealth (measured in units of output) increases. Recall that both consumption and leisure are assumed to be normal goods. The logic of the model therefore implies that the demand for both consumption and leisure should rise along with wealth with the increase in leisure coming at the expense of labor. This is called the wealth effect. Both of these effects work in the same direction for consumption (which is why consumer demand must rise). However, these effects work in opposite directions for labor supply (or the demand for leisure). The ultimate effect on labor supply evidently depends on which effect dominates; i.e., see Figure 2.3.

FIGURE 2.3
Household Response to an Increase in the Real Wage

Since theory alone cannot be used to ascertain the effect of wage changes on labor supply, the issue becomes an empirical one. As it turns out, the prevailing empirical evidence suggests that labor supply responds positively to an increase in the real wage (although, there is some debate as to how strong this effect is quantitatively). In what follows then, let us assume that the substitution effect dominates the wealth effect.
2.3 The Business Sector

The representative firm operates a production technology that utilizes labor and capital to generate output (in the form of consumer goods and services). To make things simple, assume that there is no capital, so that labor is the only factor of production. The prevailing technology is represented by a linear production function:

\[ y = zn; \]

where \( z > 0 \) is a parameter that indexes the efficiency of the production process, and \( y \) denotes the level of output. 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). Notice that with this linear production technology, \( z \) corresponds both to the marginal product of labor \((\Delta y/\Delta n = z)\) and the average product of labor \((y/n = z)\). Henceforth, I will refer to \( z \) as ‘productivity’ and exogenous changes in \( z \) as ‘productivity shocks.’

In formulating its production and hiring choices, the representative firm is assumed to take the prevailing market wage \( w \) as given (we made the same assumption for households). If a firm hires \( n \) worker-hours, then it incurs a wage bill equal to \( wn \). The employment of \( n \) worker-hours generates \( zn \) units of output, which is then ‘sold’ to the household sector. The difference between revenue \( zn \) and cost \( wn \) constitutes profit, which is subsequently handed over to shareholders (i.e., the households) in the form of a dividend payment \( d \). The objective of each firm is to maximize shareholder value (profit):

\[ d = (z - w)n. \quad (12) \]

In mathematical terms, the choice problem facing a representative firm can be stated as follows:

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

As with the household’s problem, the firm’s choice problem has a solution. Let us denote this solution by \( n^D \) (for the firm’s desired labor input, or labor demand). The solution to this particular problem is very simple and depends only on \( (z, w) \); i.e.,

---

18 This assumption is relaxed in Appendix 2.1.
19 Remember that, since there is no ‘money’ in this model economy, firms must pay workers in the form of ‘coupons’ redeemable in the output produced by the business sector.
20 That is, workers take their coupons home and then ‘shopping’ by redeeming the coupons for merchandise.
\[ n^D = \begin{cases} 
0 & \text{if } z < w; \\
n & \text{if } z = w; \\
1 & \text{if } z > w; 
\end{cases} \]

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 (again, in the form of consumer goods and services) is simply given by \( y^S = zn^D \). With this hiring and production program in place, the firm expects to generate a profit \( d = y^S - wn^D \).

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. In short, our theory predicts that the demand for labor will be decreasing in the real wage and increasing in productivity.

### 2.3 Households and Firms Together

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 fall out of thin air—ultimately, they are determined by the collective behavior of households and firms. In other words, we view market prices as being determined by conditions of aggregate supply and aggregate demand.\(^{21}\) 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

\(^{21}\)Note that for an economy populated by representative agencies, computing aggregates is very simple. In particular, if there are \( N \) agents who choose \( x \), then the aggregate is simply given by \( Nx \). In the analysis here, we have normalized \( N = 1 \).
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 agents 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

![Pattern of Exchange Diagram](image-url)
Now let us proceed to describe the general equilibrium in more detail. From the definition of equilibrium, the \textit{equilibrium} real wage $w^*$ must satisfy the labor market clearing condition:

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

Figure 2.5 provides a diagrammatic representation of the labor market.

**FIGURE 2.5**
Equilibrium in the Labor Market

Normally, one might expect the labor demand function in Figure 2.5 to slope downward smoothly from left to right. It has this peculiar 'stepped' feature here because of the linear nature of the production technology. In Appendix 2.1, I consider a slight modification to the production technology (by introducing capital and assuming a diminishing marginal product of labor) that generates a more 'normal' looking demand function. But for present purposes, this 'stepped' function simplifies things considerably, without detracting from basic intuition.

In general, the equilibrium real wage is determined by both labor supply and demand (as in Appendix 2.1). 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.

![General Equilibrium](image)

**FIGURE 2.6**
General Equilibrium

**Exercise 2.3:** Confirm that the allocation and price system depicted in Figure 2.6 satisfy the definition of a general equilibrium.

The theory developed here makes predictions concerning the determination of output and employment. It also provides an explanation for consumer spending (in this model, it is equal to GDP) and the amount of time allocated to non-market activities (leisure). Finally, it provides an explanation for what determines the real wage. These equilibrium values \((y^*, n^*, c^*, l^*, w^*)\) constitute the model’s **endogenous variables**. The theory relates the determination of these endogenous variables to the underlying structure of the economy as summarized by the set of **exogenous variables**; i.e., preferences \((u)\) and technology \((z)\). We are now in a position to ascertain how the model’s endogenous variables are predicted to change in response to any given exogenous shock.
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 this finding?

3. Understanding Business Cycles

It is an empirical fact that productivity, as measured by $z$, tends to grow and fluctuate. Why this happens is a very interesting question, but is not our concern here. Instead, the question we are interested in answering is the following: Given that productivity fluctuates over time, how might an economy, consisting of goal-oriented individuals and firms, be expected to respond to these fluctuations? As it happens, our theory is well-equipped to answer this question (which is not to say that the answer is entirely satisfactory).

Imagine that in our model economy, productivity fluctuates randomly between three levels $z_H > z_M > z_L$ (high, medium, and low). We can use Figures 2.5 and 2.6 to deduce that these productivity shocks will generate procyclical movements in employment and wages; i.e., see Figure 2.7.

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

As I mentioned earlier, the tendency for employment and wages to move in the same direction as GDP over the cycle is a feature of the data. We can use our theory to provide us with an interpretation of this observed behavior. The interpretation offered by this model corresponds closely to the intuition that was expressed earlier. In particular, when productivity is high (relative to trend), the business sector demands more labor to exploit the high return to labor. This shift in labor demand (see Exercise 2.5) puts upward pressure on real wages, which serves to draw more individuals into the labor force. The reverse holds true when productivity is low (relative to trend).
The model developed above is far too simple in many respects. But at least it provides us with a starting point. And while the model obviously abstracts from many interesting dimensions that are likely important for understanding the real world, the economic forces it emphasizes are not completely crazy. In particular, the return to labor does appear to fluctuate; and it is not implausible to suppose that individuals and firms alter their behavior in light of how this return changes over time.

What I would like to stress here is not the model per se; but rather, the methodology we employed. Our method involved being explicit about what motivates households and firms and how they are constrained in achieving their goals. We characterized the choices being made as being individually rational in the sense that supply and demand functions are the product of optimizing behavior (agents trying to do the best they can, according to their objectives, and subject to their constraints). We viewed the interaction of agents as occurring in competitive markets. This is the methodology of modern macroeconomic theory.22

---

22 There is a branch of the literature that replaces the assumption of competitive markets with alternative specifications; e.g., monopolistically competitive markets, or search markets in which prices are determined by bilateral bargaining considerations.
3.1 Policy Implications

One of the benefits of our method is that it is ideally suited for addressing questions concerning government policy. Policy questions can be divided into two categories: positive and normative. A positive question is concerned with prediction; i.e., how does policy X affect variable Y? A normative question is concerned with welfare; i.e., how should government policy be designed to improve the economic well-being of people? I will address some positive questions in the next chapter; here I am interested in studying our model’s normative implications.

Many people have a view that the business cycle is a ‘bad’ thing. This view is understandable for economic downturns, since a recession is characterized by declining incomes, wages, and employment. But this view is also expressed often during periods of economic expansion; one often hears, for example, how an economy is sometimes in danger of ‘overheating.’ Since we (in the developed world, at least) operate primarily in a market-based economy, there is a widespread perception that these recurring boom and bust episodes are somehow inherently linked to the market system itself. Evidently, the market is like a wild beast—useful for some purposes—but only if placed in shackles and guided by the hand of a skilled trainer. A natural candidate for the role of trainer is the government; the shackles often recommended take the form of stabilization policies.

Is there a role for stabilization policy in our model economy? From Figure 2.7, we can see clearly that our representative household is hurt by a recession (lower indifference curve). On the other hand, our model suggests that welfare increases during a cyclical boom (higher indifference curve). Is there any way to judge whether these equilibrium allocations are in some sense inefficient? A natural efficiency criterion is the concept of Pareto optimality. An allocation is said to be Pareto optimal if it is impossible to find a feasible allocation that improves the welfare of some person without harming the welfare of others. So, another way to approach this question is to ask whether the economy’s general equilibrium (as depicted in Figure 2.7) is Pareto optimal. In other words, could a benevolent government (one that works in the interest of our representative household) do any better (according to our Pareto criterion) than what a competitive market delivers?

As it turns out, the equilibrium of our model economy is Pareto optimal. This is not a general result, but it happens to hold true here (and can continue to hold in much more complicated environments). The policy implication is rather startling: There is no role for a government stabilization policy. In other words, even a benevolent government would choose to vary output and employment in accordance with Figure 2.7.\textsuperscript{23} Based on this model, one might

\textsuperscript{23}The way one can prove this result is to formulate the choice problem of a benevolent government:

\[\max u(c, l) \text{ subject to: } c = y = z(1 - l).\]
(mischievously and provocatively) proclaim that business cycles are optimal.

In fact, we can go further and state that—far from improving the welfare of households—government stabilization policies are likely to do the exact opposite. To show this formally, consider the following example. Imagine that the economy initially begins at point A in Figure 2.8. Imagine further that the economy then experiences a negative productivity shock (brought about, for example, by an extremely harsh winter). Natural market forces result in a recession (a decline in output, employment, and the real wage); i.e., the new equilibrium is given by point B. The movement from allocation A to B is clearly associated with a decline in welfare. But can anything be done to improve matters?

One policy that the government may consider is to stabilize the level of employment at its initial level (i.e., the level associated with point A in Figure 2.8). But as the government has no control over the decline in productivity,...

The solution to this problem corresponds to the equilibrium allocation.

24 Having worked for many years in the construction sector, I can attest first-hand to the effect that winter has on labor productivity!

25 One way the government could do this is to offer firms a wage subsidy that is financed via a...
allocation A remains infeasible. The best the government can do here is the
allocation given by point C. On the surface, such a policy may be deemed a
success. After all, note that the level of GDP and employment is higher at
point C relative to point B. But upon closer inspection, we see that (in this
model, at least) the level of welfare is lower at point C.

What is going on here? The intuition is very simple. We hypothesized that
the economy is subject to exogenous shocks to labor productivity. Individuals
may not like the fact that productivity fluctuates, but given that it does, they
respond to the implied changes in incentives (i.e., the return to work vis-à-vis
other activities) in a natural and perfectly understandable manner. In particu-
lar, it makes perfect sense for firms to scale back their demand for labor when
labor productivity falls (and expand their desired workforce when productivity
rises). Likewise, it makes perfect sense for households to adjust the way they
allocate time across competing activities when the relative return to different
activities changes. In our model, individuals respond in precisely the correct
way to changing incentives, so that collectively, individual choices result in
a socially desirable outcome. There is no need for a government to alter the
behavior of households or firms in any way.

How seriously should one take this conclusion? Personally, I believe that it
should be taken very seriously by anyone who advocates the desirability of sta-
bilization policies. This is not to say that I believe the model or its conclusions
are necessarily correct. There may very well be some role for a well-designed
government stabilization policy. What I am suggesting here is that it is not ob-
vious why or under what circumstances stabilization policy is desirable. The
simple observation that economic aggregates fluctuate and that individuals are
made worse off during a recession is not sufficient evidence to justify government
intervention (the model developed above makes this point very clear).

Because government intervention invariably implies some form of coercion, it
is incumbent upon those who advocate interventionist policies to explain (to all
of us who treasure individual freedom) why such intervention is necessary. The
value of being precise in our modeling of behavior is that it forces us to think
more deeply about the ultimate rationale for government policy. This deeper
understanding will presumably translate into better policy advice concerning
the exact nature of a well-designed government intervention.

4. Uncertainty and Expectations

It seems obvious that many—if not most—decisions must be made in the con-
text of some uncertainty over how the future will unfold. You attended class
today because you expected some return from doing so. At some point in
the future, you will realize an actual return that either meets, exceeds, or falls

lump-sum tax on households. Another alternative is for the government to simply 'command'
workers to continue working as hard as before.
short of your expectations. Likewise, in the business sector, firms must decide on how many workers to hire, or much to invest, prior to knowing the exact outcome of such decisions. Similar considerations are at work in other sectors of the economy.

The question of how expectations are formed and what role they play in determining business cycles presents a number of challenges for theorists and empiricists alike. In this section, I touch on some of the issues involved surrounding the question of expectations.

4.1 Decision-Making Under Uncertainty

Let us begin by modifying our simple model above in a way that takes uncertainty (over the return to labor) more seriously. To this end, imagine that \( z \) can take one of two values: \( z = z_H \) or \( z = z_L \) with \( z_H > z_L \) (i.e., productivity is either ‘high’ or ‘low’). Let \( p(z, s) \) denote the probability of \( z \) occurring, conditional on receiving some information \( s \) (a signal that is correlated with productivity). This signal takes one of two values: \( s = g \) or \( s = b \) (i.e., the signal is either ‘good’ or ‘bad’). Assume that a good signal implies that \( z_H \) is more likely than \( z_L \); and assume that the opposite is true with a bad signal. Mathematically, we can state this assumption as:

\[
p(z_H, g) > p(z_L, g); \quad p(z_H, b) < p(z_L, b).
\]

Let me now be explicit about the structure of information and the timing of events. Assume that a period begins with the arrival of some information \( s \). Next, prior to the realization of actual productivity, households and firms make their employment choices. Finally, the actual level of productivity \( z \) is determined. At this final stage, production and consumption takes place given the actual \( z \), and given the level of employment as determined in the previous stage.

The probability structure \( p(z, s) \) is assumed to be known by all agents in the economy. Another way of stating this is that decision-makers understand the fundamentals governing the random productivity parameter. These fundamentals are viewed as being determined by God or nature (i.e., they are exogenous). In other words, for better or worse, this is just the way things are: the world is an uncertain place and people must somehow cope with this fact of life.

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

\[ z^e(g) = p(z_H, g)z_H + p(z_L, g)z_L; \]
\[ z^e(b) = p(z_H, b)z_H + p(z_L, b)z_L. \]

Given the probability structure described in (13), 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 the model, firms always earn (in equilibrium) zero profits—both in expected and actual terms (before and after uncertainty is resolved). What this implies is that the expected wage must be given by \( w^e(s) = z^e(s) \), with the actual wage given by \( w^*(z) = z \). Of course, given the timing of events, 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.,

\[ MRS(e^e(s), 1 - n^*(s)) = w^e(s); \]
\[ e^e(s) = w^e(s)n^*(s). \]

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). In particular, actual GDP is given by \( y^*(z, s) = zn^*(s) = e^e(z, s) \).

Exercise 2.6: Using a diagram similar to Figure 2.7, draw four budget lines associated with the (actual and expected) productivity levels \( z_H > z^e(g) > z^e(b) > z_L \). Next, draw two indifference curves tangent to the budget lines associated with \( z^e(g) \) and \( z^e(b) \); these tangencies are determined by the equilibrium condition (15). From these tangencies, determine the actual levels of employment \( n^*(s) \) and the expected level of GDP \( y^e(s) \). Finally, draw two indifference curves that lie on the budget lines associated with \( z_H \) and \( z_L \) (note that these curves will not be tangent to the budget lines). Depict the actual levels of GDP \( y^*(z, s) \).

\[ \text{Instructors: For this mathematical characterization to hold exactly, we must assume that } \text{certainty equivalence holds (e.g., assume that preferences are quadratic).} \]
Exercise 2.7: Using the diagram you drew for the previous exercise, would you say that there is any sense in which individuals make ‘mistakes’ in this model? If your answer is no, then explain why. If you answer is yes, then explain what individuals (or a government) might do to correct these mistakes.

Exercise 2.8: Note that in this economy, it is possible that Nature produces a sequence of actual productivity realizations that are identical over extended periods of time. At the same time, it is possible that the signals people receive fluctuate randomly. Over such a sample period, the economy will display cyclical fluctuations in GDP and employment, even though actual productivity remains constant. Are such fluctuations ‘inefficient’ in any sense? Explain.

In the model developed here, expectations play an important role in determining resource allocation. When people are optimistic, the economy is likely to boom; when people are pessimistic, the economy is likely to bust. However, note that expectations themselves are not the cause of these boom-bust cycles. As with the simple model developed earlier, the business cycle is the product of optimal decisions in reaction to changing fundamentals. In other words, if individuals are optimistic, it is because they have good reason to be so (i.e., they have received information that leads them to revise upward their estimate of productivity). Likewise, a sudden wave of pessimism is not the product of some unexplained psychological depression; instead it is the product of a rational expectation in response to ‘bad news.’ Expectations in this model simply reflect the changing nature of economic fundamentals (information).

4.2 Animal Spirits

The nature of expectations expressed in the previous section does not appear to reflect the view held by numerous commentators and policymakers (including central bankers). The prevailing view appears to one in which psychology plays a predominant role in the formation of expectations. According to this view, exogenous (i.e., unexplained) changes in expectations themselves constitute an important source of shocks for an economy. These expectation shocks are sometimes called animal spirits—a colorful phrase coined by the famous John Maynard Keynes.27

There are a couple of ways in which to formalize this notion of animal spirits. One way is to simply assume that expectations are irrational (i.e., the product of random changes in psychological sentiment that bear no obvious relation to an economy’s underlying fundamentals). One crude way to model this is to assume that while actual productivity \( (z) \) remains constant over time, expected productivity \( (z^e) \) fluctuates for no apparent reason. Imagine then that corporate executives wake up one morning feeling ‘optimistic’ (i.e., \( z^e > z \)).

---

The expected return to production is high, which stimulates the demand for labor. The result is an expansion in employment. Unfortunately, since actual productivity is in fact given by \( z \), these high expectations invariably turn out to be unfounded. Nevertheless, the end result is a level of employment and GDP that is higher than ‘normal.’ The opposite holds true if corporate executives instead wake up in the morning feeling ‘pessimistic’ (i.e., \( z^e < z \)).

The mood swings described above are irrational because expectations do not conform with the economy’s underlying fundamentals. As a rule, (academic) economists do not take this view seriously, since it is hard to imagine that corporate executives (and other economic actors) would continue to make the same systematic errors time after time when observing no change in fundamentals. Nevertheless, such a view appears common in many circles. For example, we are often warned in the financial pages of newspapers that the economy is in danger of overheating. Even the famed Federal Reserve Chairman Alan Greenspan has, in the past, warned us of irrational exuberance in the marketplace.

What are we to make of this? While the idea of irrational expectations seems crazy to an economist, the concept does appear to offer us an insight as to what motivates many policy recommendations. In particular, if private sector decisions are influenced by a degree of irrational mood swings, then one can plainly see the benefit of a well-designed stabilization policy. In the context of the model described above, the government should act to stabilize output and employment at their fundamental levels. Of course, this policy prescription presumes that government decision-makers are not prone to forming irrational expectations. Exactly how it is that the government has the presence of mind to view things clearly, while private decision-makers do not, is never fully explained. Perhaps the answer is too embarrassing to contemplate.

Exercise 2.9: On a diagram, draw an equilibrium budget line associated with a fundamental productivity level \( z \). Now draw two more budget lines associated with an optimistic and pessimistic expectation over productivity. Assume that expectations fluctuate exogenously between these two levels. Depict the actual and expected levels of economic activity. Demonstrate how a policy that could somehow stabilize output and employment at their fundamental values would improve economic welfare.

What I have just described is just one way to think about the concept of animal spirits. As I have alluded above, this treatment is viewed skeptically by (academic) economists, as it relies on the notion of irrational expectations. There is, however, a more sophisticated way in which to formalize the concept of animal spirits. This more sophisticated treatment does not abandon the assumption of rational expectations. According to this view, expectation shocks can become self-fulfilling prophecies.

A self-fulfilling prophecy 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 consistent 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. 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 we mean by coordination failure).

This basic idea can be modeled formally in the following way. First, we need some notation to distinguish between aggregate (average) variables and individual variables. This is important because we will want to characterize optimal individual behavior conditional on the behavior of the aggregate. To this end, let aggregate (average) variables be denoted in bold font.
Now, imagine that the economy’s production technology takes the following form:

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

where \( n_C \) denotes a ‘critical’ level of 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 actual 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 (\( z \)). Now, imagine that households wake up in the morning expecting some level of \( n \). Associated with this \( n \) is a productivity (real wage) level \( z = z_H \) or \( z = z_L \) (depending on whether \( n \) is above or below \( n_C \)). Let us emphasize this dependence of \( z \) on \( n \) by writing \( z(n) \). So given some forecasted \( n \), an individual’s optimal labor supply \( n \) must satisfy the usual tangency condition:

\[ MRS(z(n)n, 1 - n) = z(n); \]

where here I have utilized the fact that \( c = z(n)n \).

In general, individual and average behavior can differ. However, in equilibrium, individual and average behavior must correspond. Thus, having characterized optimal individual behavior given aggregate behavior, we can now impose the equilibrium condition \( n = n \). The equilibrium level of employment \( n^* \) must therefore satisfy:

\[ MRS(z(n^*)n^*, 1 - n^*) = z(n^*). \]

Associated with \( n^* \) is an equilibrium level of GDP \( y^* = z(n^*)n^* \) and real wage \( w^* = z(n^*) \).

What is interesting about this model economy is that it can display multiple equilibria, with each equilibrium depending solely on initial expectations (that become self-fulfilling prophesies). In the present context, there is a possibility of a high-level and low-level equilibrium, each of which is characterized (respectively) by:

\[ MRS(z_H n_H^*, 1 - n_H^*) = z_H; \]
\[ MRS(z_L n_L^*, 1 - n_L^*) = z_L; \]

where \( n_L^* < n_C < n_H^* \). These two possible outcomes are displayed in Figure 2.9 as points A and B.
Notice that both points A and B are consistent with a rational expectations equilibrium. In the 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.’

**Exercise 2.10:** Consider Figure 2.9. 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 A.

5. Summary

One 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 (or other exogenous shocks to productivity). 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, 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. According to 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 may or 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 prophecies. In either of these cases, there is a clear role for government stabilization policies. But whether the government can, as a practical matter, identify an ‘irrational’ expectation or the existence of a self-fulfilling prophesy remains an
open question.
Problems

1. You are asked to write a brief newspaper column explaining the nature 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. Write the article in a way that my mother might understand. Limit yourself to 800 words or less.

2. Give two reasons why it is important for an 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)?
Appendix 2.1: 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 linear form \( y = zn \).

The model is easily extended to include two (or more) 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:

\[
MPL(n, z) = \theta zn^{\theta-1}.
\]

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:

\[
\text{Choose (}n\text{) to maximize } d = zn^\theta - wn.
\]

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 smoothly decreasing in \( w \) (unlike the ‘step’ function in the chapter). 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 household’s choice problem remains as before, except that now households have two sources of income (wage and dividend income). Given some arbitrary market wage \( w \) and dividend income \( d \), 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. The model’s general equilibrium is depicted as Point A in Figure 2.10.

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 2.10
General Equilibrium

Equilibrium Budget Line
Production Possibilities Frontier

$c^* = y^*$

Slope $= -w^* = -MPL = MRS$

0

$d^*$

$l^*$

1.0
CHAPTER 3

Government Spending and Taxation

1. Introduction

In this chapter, I extend the simple static model developed in Chapter 2 to include a government sector that demands some fraction of the economy’s output. This is not a chapter that explains why governments tax and spend; here we take such behavior as exogenous (e.g., determined by political factors beyond the scope of the present analysis). Instead, the theory developed here is designed to explain how one might expect output and employment to react to exogenous changes in fiscal policy (the spending and taxation policy put in place by the legislative branch of the government).

In what follows, I continue to work with a static model, so that the government will be forced to 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.

2. Government Purchases

Let us begin by assuming that the government sector ‘demands’ \( g \) units of output, where \( g \) is exogenous (and takes the form of consumer goods and services). 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 production technology as the private sector, then either approach will yield identical results.

Since government purchases (or production) of output are typically distributed in some manner across households, we need to address the question of how individuals in the household sector value \( g \). The answer to this question will depend on the nature of the purchases made by the government. In reality, the government allocates \( g \) to a wide variety of uses, ranging from outright waste to the delivery of an assortment of valuable goods and services. In what follows, I am going to assume that household preferences can be represented by a utility function of the following form:

\[
u(c, l) + \lambda v(g),
\]  

(16)
where $\lambda \geq 0$ is a preference parameter and $v$ is an increasing and concave function. For this specification, the value of $\lambda$ plays no role in determining the positive implications of our theory (i.e., the model’s predictions concerning the effect of fiscal policy on output and employment will be independent of this preference parameter). Because of this, I invoke the traditional assumption that $\lambda = 0$ (i.e., that household’s do not value government purchases). In terms of a concrete example, imagine that $g$ constitutes goods and services that are allocated for the purpose of engaging in an unpopular war.

The question that concerns us here is the following: How does an exogenous increase in government spending (an expansionary fiscal policy) affect output and employment? As it turns out, the answer to this question depends critically on the nature of the tax instrument used to finance the expansion in government purchases. In what follows, I consider two such instruments. The first is a lump-sum tax (also referred to as a head-tax). Because lump-sum taxes are rarely used in practice, this case is primarily hypothetical; nevertheless, it serves as a useful benchmark since lump-sum taxes work exclusively through their effect on household wealth (a force that is likely to be present in any type of tax). The second is an income-tax. Studying this type of tax is interesting because it is realistic and has additional interesting effects on incentives.

3. Expansionary Fiscal Policy: Lump-Sum Tax

Imagine that government spending is financed with a lump-sum tax $\tau$ on households. The key feature of a lump-sum tax is that the amount of tax paid by the household in no way depends on household behavior. What this means is that the household has no incentive to change its behavior in an attempt to escape its tax burden. Since a lump-sum tax does not distort incentives, it is called non-distortionary.

The key restriction on government behavior is given by the government budget constraint (GBC). That is, if the government desires $g$ units of output, it must levy a tax $\tau$ on the household sector sufficient to cover its desired expenditure. Mathematically, what this implies is:

$$\tau = g.$$  
(17)

To investigate how the representative household is affected by this fiscal policy, we have to reconsider its choice problem. Household preferences are given by $u(c, l)$; since here I have assumed $\lambda = 0$. The household’s budget constraint, however, must be modified to reflect its tax obligation; i.e., $c = w + d - \tau$. Substituting the time-constraint $n + l = 1$ into the budget constraint allows us

\[\text{normative}\]

\[\text{optimal level}\]

\[g^* > 0\]

\[\text{government spending}\]
to rewrite the budget constraint as: $c = w - wl + d - \tau$. The choice problem facing the representative household is therefore is given by:

$$c = w - wl + d - \tau.$$

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

**Exercise 3.1:** Assume that $d = 0$. For an arbitrary configuration of $(w, \tau)$, depict the solution to the household’s choice problem diagrammatically. Note that the following two conditions must hold: $MRS(c, l) = w$; and $c = w - wl - \tau$.

Having characterized optimal behavior on the part of households, the next step involves examining the behavior of the business sector. As it turns out, there is not much work to do here since the choice problem for a representative firm remains exactly the same as in Chapter 2. In fact, one can use the reasoning developed there to conclude—in this simple model, at least—that this fiscal policy will have no effect on the equilibrium real wage or profits (verify this fact as an exercise). Accordingly, we can conclude that $(w^*, d^*) = (z, 0)$.

Let’s see what we have so far. Optimal behavior on the part of households implies $MRS(c, l) = w$ and $c = w - wl + d - \tau$. In equilibrium, we know that $w^* = z$ and $d^* = 0$. Furthermore, we know that the government budget constraint (17) must also hold in equilibrium. Gathering these restrictions together, we deduce that equilibrium $(c^*, l^*)$ must be characterized by the following two equations:

$$MRS(c^*, l^*) = z;$$
$$c^* + zl^* = z - g. \quad (18)$$

The equilibrium level of employment can then be calculated as $n^* = 1 - l^*$; with the equilibrium level of GDP given by $c^* + g = y^* = zn^*$. Notice that the model studied in Chapter 2 is just the special case for which $g = 0$.

**Exercise 3.2:** Given preferences such that $MRS(c, l) = c/l$, use (18) to solve for $(y^*, n^*)$ as a function of $g$.

Since $g$ is a parameter (an exogenous variable), the equilibrium level of output and employment $(y^*, n^*)$ will, in general, depend on $g$. One way to investigate this dependence is through the use of algebra, as in Exercise 3.2.

---

29 This will not be the case for the more general model studied in Appendix 2.1. Nevertheless, the general conclusions arrived at here (regarding the effects of government spending on output and employment) continue to hold in even the more general model.
This approach requires us to specify an explicit mathematical form for the MRS. But we can also investigate this dependence by way of a diagram (exploiting the fact that both consumption and leisure are normal goods). To see how this can be done, consider the following.

Imagine that initially, $g = 0$. Then point A in Figure 3.1 depicts the equilibrium allocation. Notice that, as in Chapter 2, the equilibrium budget line in this case corresponds to the economy’s production possibilities frontier (PPF).30 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 households, the budget constraint moves downward in a parallel manner (leaving the PPF in its original position).31 If consumption and leisure are normal goods, then the new equilibrium is given by point B in Figure 3.1.

Thus, according to this theory, an expansionary fiscal policy financed by a
lump-sum tax induces an increase in output and employment. However, note that the policy also induces a decline in consumer spending. From the income-expenditure identity (see Chapter 1), we know that $c^* + g \equiv y^*$ in this model economy. While an increase in $g$ results in an increase in $y^*$, it also appears to partially crowd out private sector spending (so that $c^*$ falls, but by less than the increase in public sector spending, $g$).

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 households declines, they naturally demand less consumption and leisure (assuming that these are both normal goods). To put things another way, the tax on households compels them to work harder so that they might mitigate (but not eliminate) the impending decline in disposable income. This increase in the supply of labor is what leads to the expansion in GDP.

Note that while this expansionary fiscal policy causes GDP to rise, it makes our representative household worse off (the indifference curve falls to a lower level). Of course, negative effect on welfare is largely an artifact of our assumption that households do not value government spending. Nevertheless, there is an important lesson here:

\begin{center}
\text{Do not confuse GDP with economic welfare!}
\end{center}


Now, I know what some of you must be thinking here: Did we really have to go through all this trouble to demonstrate that an increase in $g$ leads to an increase in $y$? Does this not follow immediately from the income-expenditure identity $y \equiv c + g$, which we know (from Chapter 1) must always hold true? The answers to these questions are yes and no, respectively.\footnote{As well, ask yourself how you would have deduced the effects of a change in $g$ on consumption and welfare by appealing solely to the income-expenditure identity.} To demonstrate this, let me give you an example where: [1] the income-expenditure identity holds; and [2] an increase in $g$ leads to a decrease in $y$.

Consider a model economy as described earlier with an exogenous level of government purchases $g$. But instead of a lump-sum tax, suppose that the government must finance this expenditure with an income tax. If we let $0 \leq \tau \leq 1$ now denote the tax rate, then the total tax collected is given by $\tau(wn + d)$. Hence, the government budget constraint now takes the form:

$$\tau(wn + d) = g.$$  \hspace{1cm} \hspace{1cm} (19)
Let us now examine how the income tax alters the household’s budget constraint. The household’s after-tax income is now given by \((1 - \tau)(wn + d)\), so that the budget constraint is given by \(c = (1 - \tau)(wn + d)\). If we substitute in the time constraint, then the budget constraint can alternatively be written as:

\[
\begin{align*}
  c & = (1 - \tau)(w(1 - l) + d); \\
  & = (1 - \tau)(w + d) - (1 - \tau)wl.
\end{align*}
\]

**Exercise 3.3:** Depict the budget constraint on a diagram. Note that the intercept on the y-axis is given by \((1 - \tau)(w + d)\) and that the slope of the budget line is given by \(- (1 - \tau)w\). Draw an indifference curve that is tangent to this budget constraint. Note that the solution to the household’s choice problem now satisfies: \(MRS(c, l) = (1 - \tau)w\) and \(c + (1 - \tau)wl = (1 - \tau)(w + d)\).

The key observation here is that the slope of the budget constraint is now given by the after-tax wage \((1 - \tau)w\) rather than the gross wage \(w\) (as was the case with the lump-sum tax). From an economic standpoint, what this means is that an income tax distorts the relative price of output vis-à-vis leisure (with an increase in \(\tau\) making leisure relatively more expensive). This distortion has the effect of altering incentives for the household. In particular, they will be motivated (by the substitution effect) to substitute away from the more expensive commodity into the cheaper one. For this reason, an income tax constitutes an example of what economists call a distortionary tax.

Using a logic that should now be familiar, we can deduce that the equilibrium gross real wage and profits will once again be given by \((w^*, d^*) = (z, 0)\). Combining this information with the optimality conditions in Exercise 3.3, together with the government budget constraint (19), the equilibrium for this economy is the solution to:

\[
\begin{align*}
  MRS(c^*, l^*) & = (1 - \tau^*)z; \\
  c^* + (1 - \tau^*)z l^* & = (1 - \tau^*)z; \\
  \tau^* z(1 - l^*) & = g.
\end{align*}
\]

**Exercise 3.4:** Combine equations (21) and (22) to show that the income-expenditure identity \(y^* = c^* + g\) holds.

Having solved equations (20), (21) and (22) for \((c^*, l^*, \tau^*)\), we can then recover the equilibrium level of employment \(n^* = 1 - l^*\) and the equilibrium level of GDP \(y^* = c^* + g = zn^*\). I can make you do the math for a given MRS, but you’ll probably hate me for it. So why don’t we make due with a diagram instead?

We want to investigate the effect of a change in government purchases on output and employment. To begin, imagine that the government initially sets
Suppose now that there is an exogenous increase in \( g \). One thing that happens for sure is that the tax rate must increase (from zero); a fact that follows from the government budget constraint. An increase in the tax rate lowers the \textit{after-tax} wage rate (leaving the gross wage unchanged), rotating the equilibrium budget line in the manner depicted in Figure 3.2. The budget line’s flatter slope implies that consumption is now more expensive (since it is now taxed implicitly through the income tax) relative to leisure (which is not taxed). Thus, there is a substitution effect that induces each household to demand more leisure (supply less labor). Stated more simply: since an increase in the income tax reduces the after-tax return to labor, households are induced to substitute way from labor (and hence consumption) into leisure.

We are not quite done yet since an increase in the tax rate also has a negative wealth effect. In particular, lower household wealth implies a decrease in the demand for all normal goods (consumption and leisure). A reduction in the demand for leisure implies an increase in the supply of labor (this effect was also present in the case of the lump-sum tax). Thus, note that as for the case of an exogenous wage change (studied in Chapter 2), an increase in the tax rate has both a substitution and wealth effect. For consumption demand, both of these effects work in the same direction so that consumption must fall. These two forces, however, work in opposite directions for leisure. Which effect dominates depends on the nature of preferences. In Figure 3.2, I depict the case in which the substitution effect dominates (so that employment decreases); i.e., the new equilibrium is given by point B. Note that the decline in employment results in a decline in output.
Observe that it is still true for this economy that \( y \equiv c + g \). 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, this relation cannot be used to make predictive statements. This is because the income-expenditure identity is not a theory; it is simply an accounting relation.

Whether an exogenous increase in government purchases stimulates GDP and employment depends on how the added expenditure is financed.

4. Government Transfers

What is a government transfer? A transfer occurs when the government takes money (a claim to output) from person A and transfers it to person B.
Most often, person A and B are two different people. But frequently, they are the same person. For example, the government regularly takes a part of my salary in the way of a tax—and then gives it back to me in the form of a research grant! Similarly, the government collects taxes from households—and then transfers some of it back in the form of tax credits. Not everyone comes out even in this exchange, so that there is invariably some redistribution of purchasing power across households. The act of robbing Peter to pay Paul (and Peter!) constitutes an important form of government activity.33

It is important to note that while transfers constitute a form of government spending, they are not counted as a part of the GDP. Total government spending constitutes the sum of both government purchases of output and transfer payments. While government purchases (what I have labelled $g$ in the analysis above) are counted in the GDP, government transfer payments are not. The reason for this is simple: a transfer constitutes a reallocation of a given amount of output across agents; it does not therefore contribute to aggregate output. For example, suppose that you grow an apple (so that the GDP increases by the value of this apple). Imagine now that the government takes this apple away from you and gives it to someone else. Clearly, it would be wrong here to measure the GDP as increasing by two apples.

Some notation here may be helpful to clarify things. Let $T$ denote the total taxes collected by the government and let $S$ (for subsidy) denote the total transfers made by the government. As before, we can denote total government purchases by $g$. Then total government spending is given by $g + S$ and total government revenue is given by $T$. In the context of our model then, the government budget constraint can be written as:

$$g + S = T.$$  

Alternatively, the government budget constraint can be written as:

$$g = T - S,$$

where $(T - S)$ represents tax revenue net of transfers.

From an empirical standpoint, government transfer activity is important. In many developed economies, government purchases as a ratio of GDP is in the order of 20%. On the other hand, government transfers as a ratio of GDP are typically in the order of 30%. What this means is that for many countries (including Canada), total government taxes as a ratio of GDP is typically in the order of 50%! In other words, half of all the income generated in an economy is taxed, with 20% used to finance government purchases and the remaining 30% handed back to citizens in one form another.

It’s time for another one of those important lessons. I have remarked above that government transfers are not counted as part of an economy’s GDP. This

---

33This form of government activity reminds me of a radio ad I once heard from a bar (Joe Kool’s) in London (Ontario) which proclaimed: “Come to Joe Kool’s...where we screw the other guy...and pass the savings on to you!”
much is true. However, this does not mean that a government transfer policy has no effect on an economy’s GDP! Let us turn back to theory now to shed some light on this.

In the model developed above, we considered a situation where $g > 0$ and (implicitly) $S = 0$. I would now like to reverse this by assuming that $g = 0$ and $S > 0$. From the government budget constraint, what this implies is that $T = S$; i.e., all taxes collected by the government are paid out in the form of transfers (so that none of it is used to finance purchases). I will continue to use the assumption of a representative household, so that the government takes money away and gives it back in a way that does not redistribute income. As I alluded to above, this is not an entirely crazy assumption. But more importantly, the basic intuition continues to apply in a more realistic setting where households differ from each other.

In the absence of any government policy, we know that the equilibrium allocation $(y^*, l^*)$ must satisfy $MRS(y^*, l^*) = z$ and $y^* = z(1 - l^*) = c^*$. This allocation is depicted as point A in Figure 3.3.

Now, imagine that the government desires to implement a tax-transfer policy. This policy works as follows. Assume that the government distributes a lump-sum transfer of output $a$ to the representative household. This transfer is financed with an income tax $0 \leq \tau \leq 1$. Thus, in the context of our model, the government budget constraint can be written as follows:

$$a = \tau z(1 - l),$$

where here, I have invoked the equilibrium conditions $w^* = z$ and $d^* = 0$.

Given some tax-transfer policy $(\tau, a)$, the household’s budget constraint is given by:

$$c = (1 - \tau)z(1 - l) + a.$$  \hspace{1cm} (24)

The slope of this budget constraint is given by $- (1 - \tau)z$, so that optimal behavior on the part of the household implies $MRS(c, l) = (1 - \tau)z$. This implies that the new indifference curve must be tangent to the budget constraint given by equation (24); see point B in Figure 3.3.

Notice that in Figure 3.3, the new allocation $(y^0, l^0)$ is depicted as lying on top of the PPF. How can we be sure of this? The way to demonstrate this is to combine equations (23) and (24), together with the fact that $c^0 = y^0$ to derive:

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

We have just proved that the new equilibrium allocation must lie on the PPF.

But optimal behavior on the part of the household also implies that the new equilibrium allocation must lie on the budget line (24), so that:

$$y^0 = (1 - \tau)z(1 - l^0) + a.$$  

As you can see, point B satisfies this condition as well.
The only thing left to demonstrate is whether the new allocation lies to the ‘right’ or ‘left’ of point A. I claim that the new equilibrium must be at a point to the ‘right’ of point A. How do I know this? At point A, I know that the following is true: \( MRS(y^*, l^*) = z \). At point B, on the other hand, the following must be true: \( MRS(y^0, l^0) = (1 - \tau)z \). This implies that \( MRS(y^0, l^0) < MRS(y^*, l^*) \), so that the indifference curve at point B must be ‘flatter’ than the one at point A. This can only occur if point B lies to the right of point A. Therefore, the new equilibrium is characterized by an allocation as depicted by point B in Figure 3.3.

In this example, the income-expenditure identity is given by \( c = y \) (since \( g = 0 \)). Government expenditures are positive (i.e., \( a > 0 \)), but these expenditures take the form of transfers and hence are not counted as a part of GDP. Despite this fact, we see from the model developed above that a government transfer policy is likely to affect the level of GDP (output and employment are both
The fact that government spending on transfers is not counted as a part of GDP does not imply that a government transfer policy has no effect on GDP!

In our model, there are two effects that work to reduce employment (and hence output). First, as the government increases the ‘generosity’ of the lump-sum transfer, households can afford to consume more leisure (and hence, reduce labor supply). Second, the income tax reduces the return to work, which further serves to reduce labor supply.

In the context of our example above, which 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.
Problems

1. 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 3.4 for the case of the United States during World War II. Are the patterns of economic activity in Figure 3.4 consistent with our theory? Explain. Hint: draw 'trend' lines through the data in the diagrams below.

![Figure 3.4](Gross Domestic Product and Employment in the United States During WWII)

2. Does a distortionary tax necessarily imply an adverse effect on labor supply? Derive the labor supply function for preferences given by \( MRS = \frac{c}{l} \) and a budget constraint \( c = (1 - \tau)w(1 - l) \). How does an increase in \( \tau \) affect labor supply here? Explain. Suppose instead that preferences are given by \( MRS = (c/l)^{1/2} \). How does labor supply depend on the tax rate for these preferences? Explain. (Diagrams will be helpful).
3. You work hard during the year to produce one apple. This contributes to the annual GDP by the value of the apple. Suppose that the government takes this apple away from you and redistributes it to someone else. Explain why this government policy will not affect this year’s GDP (given that you have already produced the apple). On the other hand, explain why such a policy, if it is to remain in place, is likely to reduce future GDP by one apple.

4. It is a simple matter to extend the theory above so as to endogenize the level of government purchases. For example, imagine that preferences are given by: \( u(c,l) + \lambda v(g) = \ln(c) + \lambda \ln(g) \). Note that for these preferences, the household is assumed not to value leisure, so that \( n^* = 1 \) and \( y^* = z \) (this is just for simplicity). For these preferences, we have \( MRS(c,g) = c/(\lambda g) \). The PPF is given by \( c + g = z \), so that it’s slope is given by \(-1\). For a government concerned with maximizing the welfare of the representative household, the optimal allocation \((c^*, g^*)\) is determined by \( MRS(c^*, g^*) = 1 \) and \( c^* + g^* = z \). Depict this on a diagram and solve for \((c^*, g^*)\) as a function of the parameters \((\lambda, z)\). According to this theory, an exogenous increase in productivity should bring forth an expansion in government purchases. Explain why.
1. Introduction

To this point, 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. To understand how such decisions are made, we need to develop a dynamic model.

In this chapter, I focus on the consumption-savings choice of individuals. Since any act of saving serves to reduce consumption in the present and potentially increase consumption in the future, the key decision involves how to best 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. 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.

2. A Simple Dynamic Model

2.1 Preferences and Endowments

Our model economy consists of a representative household that lives for two periods (the entire duration of the economy). Each household has 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 (like apples and oranges, or output and leisure). The
assumption that people have preferences for time-dated consumption simply re-
fects 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, assume that there is no uncertainty over how the future evolves
(this can be easily relaxed).

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 of 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 a household’s
indifference curves in the commodity space.

![Figure 4.1 Indifference Curves](image)

Each household is endowed with an **exogenous** output profile \((y_1, y_2)\), which
constitutes a point in the commodity space. Since output is exogenous, the
model constitutes an example of what is called an **endowment economy**. That is to say, output (per capita GDP) is not produced; it is simply endowed to
households by Nature.\(^{34}\) Assume further that output is **nonstorable**. In other

\(^{34}\)Alternatively, one can think of each household as being endowed with one unit of time
and a production technology \(y = zn\). If households do not value leisure, then \(n^* = 1\) and
\(y^* = z\). In other words, output is simply given by the (exogenous) level of productivity.
words, one cannot store output across time (say, in the form of inventory). This assumption turns out to be important, and will be relaxed in a later chapter.

2.2 Robinson Crusoe

Let’s pause here for a moment and talk about English literature. In the year 1719, Daniel Dafoe published his famous novel *Robinson Crusoe*, which is about a sailor shipwrecked on a deserted island. Dafoe’s novel has inspired numerous books and films, including *Gilligan’s Island* and *Castaway* (maybe you’ve seen this one—it stars Tom Hanks).

Enough with pop culture. I mention Robinson Crusoe only because his name occasionally appears in economic models as a metaphor for an environment that leaves a person with no opportunity to trade with others. The model, as I have described it so far, has this feature. That is, assume that Robinson Crusoe (a representative household) has preferences for coconuts today and coconuts tomorrow. Coconuts are nonstorable (so that if they are not eaten, they spoil). Nature has endowed Crusoe with a coconut tree that yields $y_1$ coconuts today and (is expected to yield) $y_2$ coconuts tomorrow. Mathematically, Crusoe’s choice problem can be stated very simply as:

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 consume his entire income in each period (with the level of income dictated by Nature). This solution is depicted diagrammatically in Figure 4.2.

---

what follows then, we can think of shocks to $y$ as being the result of shocks to productivity. Finally, note that since $z$ is also the marginal product of labor, $y$ will also correspond to labor earnings.

35If you do not know what a metaphor is, look it up at www.dictionary.com. A favorite of mine involves a Russian hockey scout commenting on a young Russian hockey prospect: “Ah yes...Boris...he is strong like bull...and smart like refrigerator!”
Why is this interesting? Recall that a Robinson Crusoe economy is one in which the inhabitants of an island (country) have no opportunity to trade with others (in particular, households occupying other islands—or countries). Among other things, what this means is that Robinson Crusoe cannot undertake financial transactions with foreigners; i.e., the economy is closed to international trade in goods and assets. Such a restriction may be imposed by Nature (as in the case of Robinson Crusoe), or may be imposed by government policy (as in the case of Albania, under the rule of Enver Hoxha, 1944-85). Whatever the reason, our model makes clear that an economy that is closed in this manner must live within its means on a period-by-period basis. In other words, consumption cannot exceed income at any point in time (imagine imposing such a restriction on university students!).

Exercise 4.1: The marginal propensity to consume (out of current income) is defined as $MPC \equiv \Delta c_1^D / \Delta y_1$ (the change in desired current consumer spending as a result of a ‘small’ increase in current income). Compute the MPC for Robinson Crusoe. How does $c_1^D$ depend on $y_2$ (the expected level of future income)?
2.3 An International Bond Market

In this section, I describe how Robinson Crusoe’s circumstance is altered by the presence of an international financial market. The financial market is very simple: it features only one type of financial instrument—a risk-free private debt instrument called a **bond**. A bond constitutes a **promise** (made by the issuer) to deliver something of value at some future date (to the bond-holder). Since I want to abstract from bond risk, assume that bond-issuers always keep their promises (this allows us to ignore the complications that arise when default is a possibility).

We have to be clear here about the exact nature of the bond being considered. In the real-world, most bonds constitute promises to deliver money. For example, suppose that you issue a bond for $10,000 (a student loan), which you promise to redeem (pay back) one year from now to the bond-holder (a bank). This is an example of what economists call a **nominal bond**. The $10,000 constitutes the **principal** amount of the bond. If your debt contract with the bank requires you to pay back $11,000, then the extra $1000 constitutes the **nominal interest** paid on the principal. The **nominal interest rate** in this case is 10% per annum.

In our simple model, there is no role for money. People are assumed to care only about time-dated consumption (and you can’t eat money—at least, it doesn’t taste very good or have much nutritional content). Since money is not valued, there is obviously no role for a nominal bond. However, there is a potential role for what economists call a **real bond**. A real bond constitutes a promise to deliver **output** (i.e., in the form of future consumption). For example, suppose that you are off to a party one night and ask your roommate to lend you a case of beer (current consumption). You promise to pay your roommate back the next day with a new case of beer (future consumption). The principal amount of this loan is 12 bottles of beer. The **real interest rate** in this case is 0%. Notice that no money has changed hands in this financial transaction (note too that beer is more liquid than money).³⁶

In Chapters 2 and 3, we considered a static model with two goods: output and leisure. These two goods exchanged in one market (a market in which households exchanged leisure for output). When only two goods can be exchanged, there can only be one price. In Chapters 2 and 3, this price was the real wage (the exchange rate between output and leisure). In the model studied here, there are only two goods: current consumption and future consumption. We want to think of these two goods as being exchanged in one market (a market in which households exchange current consumption for future consumption—i.e., a financial market). The relevant price in this case will be the (gross) real rate of interest (the exchange rate between current and future consumption). Let me

³⁶That last part was just a joke. But in case you are questioning the empirical relevance of this example, note that many governments in fact do issue real bonds (nominal bonds indexed to the price-level).
Imagine that our representative household has access to an international bond market. In this case, the domestic economy is said to be open to trade (in financial instruments). Let $R$ denote the (gross) real rate of interest prevailing in the bond market. For now, we view $R$ as an exogenous variable; i.e., the collective trading behavior of our domestic residents is assumed to have no effect on the interest rate prevailing in world financial markets. In this case, we are dealing with what economists call a small open economy. Assume that there is free trade in the sense that households are free to buy and sell bonds at the prevailing market interest rate.

The opportunity to buy and sell bonds implies an opportunity to save or borrow. In general terms, saving is defined as current income net of expenditure on current needs. Keep this definition in mind, since it will be useful whenever you want to think about calculating saving.

Assuming that our representative household begins time with no outstanding assets (or debt), current period income is given by the GDP. Further, since there is no government sector, expenditure on current needs is given solely by consumer spending. Thus, in the context of this model economy, we can define net domestic saving as:

$$s \equiv y_1 - c_1.$$  \hspace{1cm} (25)

Note that in a closed economy (with no investment opportunities), it must necessarily be the case that $s = 0$. In an open economy, however, it is possible to have $s > 0$ or $s < 0$. In the former case, we say that the country is a net lender; in the latter, we say that the country is a net borrower. Again, 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 (recall that a real bond represents a claim to future output).

As you all likely know from personal experience, the act of lending or borrowing has implications for your future opportunities. The same is true for a country as a whole. If a country lends $s$ units of output today, it expects to receive the principal and interest on this loan in the future. This implies $Rs$ units of future output in addition to whatever is produced domestically. In mathematical terms:

$$c_2 = y_2 + Rs.$$ \hspace{1cm} (26)

In case you missed it, note that the gross interest rate $R$ is related to the net interest rate $r$ by the equation $R = 1 + r$.

Keep in mind that equation (26) holds whether saving is positive or negative. In the case of a net borrower ($s < 0$), the quantity $Rs$ represents the principal.\footnote{The gross interest rate $R$ is related to the net interest rate $r$ by the equation $R = (1 + r)$.}
and interest that is repaid on accumulated debt. Thus, the act of borrowing today has the effect of reducing future consumption.

Before moving on, I would like to mention one more thing. In case $s \neq 0$, future income has two components. First, there is income that is produced domestically; this is the GDP $y_2$. Second, there is the income that is earned on (or paid to) financial assets (debt); this is given by $rs$ (net foreign interest income). If $s > 0$, then $rs$ represents net interest income received from foreigners; if $s < 0$, then $rs$ represents net interest income paid to foreigners. What this implies is that in an open economy, the GDP need not equal the GNP (see Chapter 1). In particular, note that the future GNP in this economy is given by $y_2 + rs$. The $s$ here represents nationally-owned assets ‘employed’ on foreign soil (a negative $s$ implies foreign-owned assets ‘employed’ on domestic soil).38

Alright, let’s move on. Take the definition of saving in equation (25) and insert it into equation (26) to derive: $c_2 = y_2 + R(y_1 - c_1)$. This equation constitutes a budget line referred to as the intertemporal budget constraint (IBC). Using a diagram depicting the commodity space, this budget line has a y-intercept equal to $Ry_1 + y_2$ and has a slope equal to $-R$; i.e., see Figure 4.3.

38 Note that since these assets are purely financial in nature (as opposed to consisting of physical capital), they contribute nothing in the way of production (which is why they do not affect the GDP).
Let me note several things of interest that are evident from Figure 4.3. The first thing to note is that the availability of an international bond market greatly expands Robinson Crusoe’s intertemporal consumption opportunities (compare Figure 4.3 with Figure 4.2 in terms of what is feasible). Second, we know from basic theory that the slope of a budget line constitutes a relative price. Evidently, the gross real rate of interest \( R \) represents the **intertemporal price of consumption**. In particular, \( R \) represents the price of current consumption relative to future consumption. An increase in \( R \) makes current consumption more expensive (makes future consumption less expensive). Alternatively, think of \( R \) as the return to saving. Since saving is used to finance future consumption, an increase in \( R \) makes future consumption easier to obtain. Third, note that the IBC passes through the endowment point \((y_1, y_2)\). Note that this must always be the case, since choosing \( s = 0 \) is always an option.

The IBC can be rearranged in the following useful way:

\[
c_1 + \frac{c_2}{R} = y_1 + \frac{y_2}{R}
\]  

(27)

The right-hand-side of the equation above is the **present value** of the economy’s GDP flow. 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 economy’s GDP flow (and is depicted by the y-intercept in Figure 4.3).

---

**Do not confuse income (GDP) with wealth. Income is a flow (a sequence of numbers); wealth is a stock (a single number that measures the value of the income flow).**

---

Now, if the right-hand-side of (27) measures the present value of the GDP flow, you can pretty well guess (correctly) that the left-hand-side represents the present value of the economy’s consumption flow. (Keep in that the principles involved here apply equally to an individual as they do to the economy as a whole).

The IBC imposes a restriction on behavior. It tells us that the consumption flow \((c_1, c_2)\) must be such that its present value does not exceed the economy’s wealth. Note that this restriction is much **weaker** than the one facing Robinson Crusoe in the absence of a bond market. Robinson Crusoe (an economy closed to international trade in financial assets) is constrained to live within his means on a period-by-period basis; i.e., \( c_1 = y_1 \) and \( c_2 = y_2 \). In contrast, with access to a bond market, an economy can potentially detach its consumption profile from its income profile so that \( c_1 < y_1 \) and \( c_2 > y_2 \) or \( c_1 > y_1 \) and \( c_2 < y_2 \) are possibilities. In other words, with access to an international financial market, an
economy is no longer restricted to live within its means on a period-by-period basis. Instead, the economy is restricted to live within its means only in an intertemporal sense (i.e., on a lifetime basis).

**Exercise 4.2:** 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?

The exercise above demonstrates that just because one can borrow (live beyond one’s current means), this does not mean that one can borrow infinite quantities. Creditors will only lend you (or countries) resources if they expect you to have the means to pay back the loan in the future. If the (gross) interest rate on a loan is \(R\), then the maximum you can borrow is \(y_2 / R\). In some sense, \(y_2\) is serving as collateral for the loan. **No collateral = No loan.** Or, to put things another way: **Loan + No Collateral = Charity.**

**Exercise 4.3:** What form of collateral do students (at least, implicitly) offer creditors (e.g., a bank or the government) when they take out a student loan? If students could default on their loans with impunity (a policy often advocated by student unions and other nut cases), what do you think would happen to the supply of student loans? Explain.

Before concluding this section, let me introduce some more terminology. Recall the income-expenditure identity \(Y = C + I + G + X - M\) (see Chapter 1). The value of exports \(X\) net of the value of imports \(M\) is called the trade balance (or net exports); let \(TB \equiv X - M\). If \(G\) takes the form of consumption (expenditure on current needs), then domestic saving is defined as \(S \equiv Y - C - G\). Combining these identities, we see that \(S = I + TB\). As there is no investment in our model economy, we have \(S = TB\). In other words, net saving is only possible by running a trade balance surplus; conversely, net borrowing is possible only by running a trade balance deficit.

**Exercise 4.4:** Use the IBC in equation (27) together with the definition of the trade balance to show that: \(TB_1 + \frac{TB_2}{R} = 0\). If a country is currently running a trade balance deficit \((TB_1 < 0)\), what does this imply about the future trade balance? Explain.

Related to the trade balance is the concept of the current account position. An economy’s current account (CA) is defined as \(CA \equiv GNP - C - G\). The only difference between these two concepts is that the trade balance is defined using GDP \((Y)\), while the current account is defined using GNP. In our model economy, the current account (for each period) is given by:

\[
CA_1 = y_1 + 0 - c_1; \\
CA_2 = y_2 + rs - c_2.
\]
Thus, it is possible, in principle at least, for a country to simultaneously run a current account surplus and a trade balance deficit (e.g., \(CA_2 > 0\) and \(TB_2 < 0\)). This would be the case, for example, for a net creditor nation whose citizens held a large quantity of foreign financial assets. Of course, the converse is also true.

The key thing to take away from this last bit is that the trade balance and the current account are both inextricably linked to the consumption-saving decisions of domestic households.

### 2.4 Consumption-Saving Behavior

Now that we have described the representative household’s preferences and constraints (IBC), we can derive the household’s optimal consumption-saving plan, conditional on a set of parameters \((y_1, y_2, R)\). In mathematical terms, the decision-problem can be stated 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 \((y_1, y_2, R)\). Once \(c_1^D\) is known, one can calculate the household’s (economy’s) desired saving function \(s^D\) from the definition of saving; i.e., \(s^D = y_1 - c_1^D\). The solution is depicted graphically in Figure 4.4 as point A.

There are two mathematical conditions that 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,

\[
\frac{d}{dI} = R; \\
\quad c_1^D + \frac{c_2^D}{R} = y_1 + \frac{y_2}{R}. \tag{28}
\]

**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 \(\beta\) can be interpreted as a ‘patience’ parameter. In particular, what would \(\beta\) be equal to for an individual who ‘doesn’t care’ about the future?

**Exercise 4.6:** For the utility function in the previous exercise, one can derive \(MRS(c_1, c_2) = c_2^D / (\beta c_1^D)\). Use the two equations in (28) to derive the consumer demand functions \(c_1^D, c_2^D\) and the desired saving function \(s^D\). How is \(c_1^D\) predicted to depend on \(\beta\)? Explain whether or not this makes sense to you.
Exercise 4.6: Using a diagram similar to Figure 4.4, depict a case in which households have a low current income, but expect a higher future income. Explain how this theory can also be used to explain why students go into debt. Is it a good thing that students are allowed to go into debt? (Use this theory to show how welfare is affected when students are prevented from taking on debt).

Figure 4.4 depicts a situation in which this economy is running a trade balance surplus in the current period (and a trade balance deficit in the future period). Whether a country runs a surplus or deficit depends on the configuration of parameters. In Figure 4.4, I have placed the endowment point in a position that implies that current GDP is significantly higher than the expected future GDP. It is not surprising then, that given the forecasted decline in GDP (falling income for domestic households), households wish to save in order to smooth their consumption spending across time. The way in which domestic households save is by purchasing foreign bonds (claims to future consumption). The way households pay for these bonds is by exporting current output—which is what accounts for the trade balance surplus in the current period.

Exercise 4.7: Consider a parameter configuration such that a small open economy is currently running a trade balance deficit. Many governments have in the past
implemented **capital controls**, for example, a legal restriction that prevent foreign agents from purchasing domestic financial assets. Presumably, the idea here is that a trade balance deficit is a 'bad thing,' and that the capital controls will serve to prevent an 'excessive' deficit from occurring. Use the theory developed here to show that while capital controls can successfully reduce (or eliminate) a trade deficit, this is likely to come at the expense of lower economic welfare for domestic agents.

3. Small Open Economy Response to Shocks

The model developed above constitutes a theory of household consumer demand (and saving). Alternatively, in the context of a small open economy, we can aggregate across households to develop a theory that explains the determination of aggregate consumer spending and the trade balance (or 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 GDP (or productivity). In particular, 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. It follows therefore that, to an outside observers, an economy’s trade balance may move for no apparent reason (when it is, in fact, responding to new **information** concerning the likely path of future GDP).

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, and how does this behavior manifest itself in terms of the trade balance (or current account)?

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

---

39 Of course, this aggregation exercise is easy for an economy populated by a representative household. But keep in mind that one can easily extend this analysis to incorporate heterogeneous households. Doing so would not significantly alter any of the conclusions highlighted in this chapter (but would render welfare analysis a little trickier).
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, beginning from this position, 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 time-dated consumptions are normal goods, 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](image)

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., $\Delta c_1^D < \Delta y_1$. Recall from a previous exercise that the ratio $\Delta c_1^D / \Delta y_1$ (for a ‘small’ $\Delta y_1$) 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 unit increase in income results in a less than one unit 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 foreign bond purchase 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 households (and hence the economy as a whole) 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) internationally, households in an economy can use the financial market to absorb the impact of transitory income shocks and in this way keep their lifetime consumption patterns relatively stable.

Access to an international financial market allows a small open economy to smooth consumption (living standards) in the face of transitory fluctuations in GDP. In this sense, the financial market serves as a type of 'shock absorber.'

Exercise 4.8: Imagine extending our 2-period model to allow for many time periods T. Consider a shock to current GDP of a given size. Explain why the response of current consumer spending is likely to grow smaller as the time horizon T is made larger.

Exercise 4.9: In a recent article, John Bluedorn 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 4.10: I began my academic career at the University of Waterloo in 1991. I recall at that time picking up a Toronto newspaper and reading an article that stated the following: “Canada has just entered a recession. Adding to our

\[40\text{www.economics.ox.ac.uk/faculty/EconDetails.asp?Detailno=173}\]
problems is the ballooning current account deficit.” Explain why the ‘ballooning’
current account deficit was 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).

3.2 Good News/Bad News

A ‘news shock’ refers to an exogenous shock to information concerning the
future, leading households to alter their expectations concerning the likely
path of future events. The particular news shock I want to consider here involves
the arrival of new information that leads households to revise upward their
forecast of future earnings. We can think of this as ‘good news.’ Everything I
say below holds in reverse for ‘bad news.’

A shock to news about future productivity does not affect current produc-
tivity; we can model this by setting $\Delta y_1 = 0$. Good news here can be modeled
as $\Delta y_2 > 0$. An example of such news could be what typically seems to happen
just before an economy emerges from recession (households become optimistic
of an impending recovery). Alternatively, one could imagine the arrival of a new
technology that is expected to improve GDP in the near future. How do our
model households 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., $\Delta c_{1D} > 0$ and
$\Delta c_{2D} > 0$. This is just the consumption-smoothing motive at work again.
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.
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 trade balance moves to a 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 serves as a type of shock-absorber for consumption.

The example portrayed in Figure 4.6 also reveals another important point. Notice that while the trade balance (and current account) 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. An important lesson to be drawn here is that we must be careful in drawing any immediate link between a country's current account position and the welfare of its residents. In particular, note that a trade balance deficit may be the product of a bad shock (e.g., a recession) or of a good shock (optimism over future prospects).
A decline in the trade balance may be the product of either good or bad shocks. There is no theoretical justification for utilizing the trade balance as a measure of economic welfare.

Exercise 4.11: 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, you live in a cardboard box and you subsisting largely on Kraft macaroni and cheese 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). Hint: collateral.

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 households 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 precede the actual rise in aggregate income (GDP).

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.
Do not confuse correlation with causality. In particular, the fact that consumer spending is correlated with future GDP does not imply that consumer spending stimulates the economy.

**Exercise 4.12:** 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 $(y_2/y_1)$.

### 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° shift of the endowment (A → 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 larger 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.
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 $\Delta c_D/\Delta y = 1.0$. In other words, 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.

**Exercise 4.13:** For preferences such that $MRS = c_2/(\beta c_1)$, derive the current period consumer demand function $c^D_1(y_1, y_2, R)$. Demonstrate that $\Delta c^D_1/\Delta y_1 < \Delta c^D_1/\Delta y$, where $\Delta y = \Delta y_1 = \Delta y_2$. Does this theoretical prediction make sense to you? To answer this, think of your how you are likely to behave under the following two circumstances. In scenario one, you arrive to class and spot a $100 bill on the seat. In scenario two, you know (or expect) that there will be a $100 bill on your seat throughout the entire semester. On the first day of class, your income is the same under both of these scenarios. I would venture to guess, however, that your spending pattern is likely to differ.
3.4 A Real Interest Rate Shock

How is an exogenous change in the real rate of interest predicted to influence behavior? As mentioned earlier, the real interest rate is an intertemporal price; i.e., it measures the relative price of output across different time periods. Any change in the market interest rate will have implications for the ability of households to substitute consumption intertemporally. As incentives for intertemporal substitution change, desired consumption and saving patterns are likely to change as well, with the corresponding implications for the trade balance (and economic welfare).

We know from basic theory that a change in prices will generally have both substitution and wealth effects. In terms of saving behavior, these two effects happen to work in the same direction for borrowers, but in opposite directions for lenders. Below, I consider each case in turn.

In our model, lenders are likely to be characterized by households with high current income and low future income (i.e., \( y_1 > y_2 \)). A classic example here would be individuals in their peak earning years who are expecting to retire in the not-too-distant future. Alternatively, we can think of a small open economy that is currently experiencing a transitory boom in GDP.

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.
In our model, borrowers are likely to be characterized by households with low current income and high future income (i.e., $y_1 < y_2$). A classic example here would be young people with their peak earning years approaching in the not-too-distant future. Alternatively, we can think of a small open economy that is currently experiencing a transitory recession.

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: Households desire 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 $c_1$ (and hence saving). 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.
4. Borrowing Constraints

The theoretical analysis so far has assumed that households (small open economies) are free to borrow or lend at the market interest rate. In reality, however, this may not always be the case. Some economists argue that households are generally free to save, they often have trouble borrowing so that, in addition to their intertemporal budget constraint, some households (small open economies) are subject to a borrowing constraint.

Before going further, let me be clear about the distinction between the IBC and a borrowing constraint. I have already remarked above that the IBC implies a constraint on borrowing. In particular, for a given interest rate $R$ and future income $y_2$, the IBC implies that a household can borrow no more than $y_2/R$ (the present value of its future income flow). Extending credit beyond this point is not an act of lending; but rather, it is an act of charity (since the household would not have the means to pay back the principal and interest on the debt).

A borrowing constraint refers to a situation in which a household has the ability to make good on a debt, but where creditors nevertheless refuse to extend a desired amount of credit below this level. One can model this by assuming that creditors are willing to use as collateral for a loan only some
fraction $0 \leq \theta \leq 1$ of future income, so that while a household has the ability to repay a loan as large as $y_2/R$, creditors will only allow a loan size as large as $\theta y_2/R$. In the case where $\theta = 0$, households cannot borrow at all (even if their IBC implies that they have the ability to repay the loan).

To demonstrate how a borrowing constraint can influence behavior, consider Figure 4.10. Imagine that the household has an endowment given by point B. In the absence of a borrowing constraint, this household would like to borrow (i.e., choose point A). Clearly, point A does not violate the household’s IBC. On the other hand, suppose (for whatever reason) that $\theta = 0$, so that creditors refuse to lend any resources at all to this household. Then the best that this household can do is to choose point B. In this case, we say that the borrowing constraint is binding.

**FIGURE 4.10**  
Borrowing Constraints

**Exercise 4.14:** Consider a household that is facing a binding borrowing constraint.  
Explain why the marginal propensity to consume out of current income is larger for this household relative to some other household that is not similarly constrained.

Do borrowing constraints exist in reality? The answer to this question is not as straightforward as one might imagine. As an empirical matter, it is
often difficult to ascertain whether a household that claims to have trouble borrowing is simply running up against its IBC or whether it is truly borrowing constrained. One problem here is that it is difficult for creditors to observe the household’s future income \( (y_2) \). In particular, a household (or country, for that matter), may claim to have a high \( y_2 \), when in fact it does not. The world is, unfortunately, full of people wanting to ‘borrow’ with little intention of ever repaying their debt.

If a borrowing constraint exists, it is likely there for an important reason: a lack of commitment on the part of the borrower. The IBC assumes that debtors can keep their promises; either willingly, or because of a legal system that enforces contractual terms. It is often the case, however, that governments impose laws that prevent creditors from enforcing their claims. A law that allows individuals to declare personal bankruptcy constitutes one such example. These laws often allow debtors to discharge (get rid of) their debt to private creditors with virtual impunity.\(^41\) If a potential debtor has the option of declaring bankruptcy, a creditor may question the level of commitment and scale back the size of the loan accordingly.\(^42\) For some reason, this is often referred to as a ‘financial market imperfection’ when, in fact, it represents the logical outcome of a government legal restriction.

The lack of commitment power is, if anything, even more severe in an international context. If your neighbor doesn’t want to pay you back for a loan, you can take him to court and have a sheriff seize his assets on your behalf. If a foreign country does not want to make good on its debt, there is very little you can do to collect your money (sending in a sheriff to seize property would be tantamount to a declaration of war). Given this fact, one might be forgiven for wondering how international financial markets operate at all. Why don’t debtor nations simply default on their foreign debt obligations?

The incentive for default can be seen plainly in Figure 4.10. That is, imagine that a small open economy initially sells bonds to foreigners, allowing it to import goods and attain point A. Presumably, foreigners only extended this credit in the first place because they expected to be repaid. But imagine now that our small open economy ‘surprises’ creditors (in the second period) by refusing to make good on its obligations. In doing so, future consumption jumps from point A to point C. This looks like a good deal for domestic agents. But if they get away with it once, do you think that they could get away with it again? Probably not (at least, not for a while). Would you be willing to lend to (purchase the bonds of) agents of a country with a track-record of default? We see then that there is some incentive to pay back your debt (even if you could default) so as to protect your reputation. A good reputation (good credit history) is valuable because it will likely grant you access to the loan market.

\(^41\)It is interesting to note that one may not typically discharge debt owing to the government in a bankruptcy proceeding. The government only allows you to discharge debts owing to private sector agents. Clever, aren’t they?

\(^42\)In effect, the bankruptcy law legally prevents a debtor from using some or all of future earnings as collateral for a loan. As I remarked earlier: no collateral = no loan.
some time in the future when you might really need it. The same principle holds true for individuals, businesses, and governments. To the extent that this reputation mechanism works well, borrowing constraints are unlikely to be a quantitatively significant problem.

5. Determination of the Real Interest Rate

To this point, we have treated the real rate of interest as an exogenous variable. 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 economy in which they operate, then their individual actions are unlikely to affect market prices. That is, from the perspective of a small decision-making unit, it is as if market prices bounce around exogenously according to some law of nature.

Nevertheless, it remains true that the real interest rate is a market price and that market prices are not determined by God or Nature; rather they are determined, in part, by the collective behavior of market participants. 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.

To think about the determination of the interest rate, let’s gather all of the world’s small open economies and consider them as one economy (the world economy). This world economy is a closed economy (unless you believe that some of us are trading with aliens). Note that while it is possible for a small open economy to run a trade balance surplus or deficit, the same cannot be true of a closed economy. As far as the world economy is concerned, the trade balances of all countries together must sum to zero.

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.
5.1 General Equilibrium in a 2-Period Endowment Economy

Consider Figure 4.4. This figure depicts an economy’s desired consumption (and saving) profile given some intertemporal pattern of GDP \((y_1, y_2)\) and given some (arbitrary) real rate of interest \(R\). In this section, I continue to view \((y_1, y_2)\) as exogenous (which is why this is called an endowment economy). But I now ask the question: What determines \(R\)?

To answer this question, we will have to reinterpret Figure 4.4 as depicting the world economy. That is, let us now interpret \((c_1^D, c_2^D)\) as the consumption profile of a representative household 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 trade balance 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 net world saving (the world trade balance) must be equal 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.

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 \(R^*\).\textsuperscript{43}

\textsuperscript{43}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 \(s_i^D \geq 0\) for \(i = 1, 2, ..., N\) as long as \(\sum_{i=1}^{N} s_i^D = 0\).
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:

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

Equation (29) tells us that, in theory, the real rate of interest is determined in part by preferences (the patience parameter $\beta$) and in part by the expected growth rate of the world economy ($y_2/y_1$).

5.2 A World Recession

Recall that we have previously modeled a recession as a transitory decline in current GDP; i.e., $\Delta y_1 < 0$ and $\Delta y_2 = 0$. If such a shock is correlated across small economies, then the world economy will experience a transitory decline in GDP as well. We can then use our general equilibrium model to help identify what sort of forces come into play in the international bond market.

115
To answer general equilibrium type questions, it is useful to proceed in two steps. First, consider the effect of the shock in a partial equilibrium setting (i.e., assuming that prices remain constant). Second, identify what must happen to prices to restore the general equilibrium.

Consider then the initial general equilibrium, given by point A in Figure 4.12. A transitory decline in GDP moves the endowment profile to point C. Imagine, for the moment, that the real interest rate remains unchanged at its initial level. Then by a standard argument, the indifference curve shifts to point B (this is just the small open economy analysis). But we see that point B is inconsistent with general equilibrium. In particular, there is an excess demand for borrowing (all countries want to borrow imported goods and services to help smooth consumption). The competition for loanable funds must drive the interest rate higher; general equilibrium is restored 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 the economic forces at work here.
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, a household (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 over time. In other words, the concept of consumer demand should also be thought of as the solution to a dynamic choice problem.

With the availability of financial markets, households (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, the level of the trade balance is related to domestic saving decisions. A trade balance surplus corresponds to the lending of output to foreigners; while the converse holds true for a trade balance deficit. Whether a country is in a surplus or deficit position reveals nothing about the welfare of domestic residents. A large increase in the trade balance deficit may, for example, be the result of either a domestic recession (lower welfare) or ‘good news’ concerning the future growth prospects of the domestic economy (higher welfare).

The real interest rate is an intertemporal price; it measures the relative price of consumption across time. For a household (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).
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](image)

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 \( MRS = \frac{c_2}{c_1} \). Show that the consumption-output ratio \( \left( \frac{c^D}{y_1} \right) \) is given by:

\[
\frac{c^D}{y_1} = \frac{1}{2} \left( 1 + R \frac{y_2}{y_1} \right).
\]

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
Dominica
Real per capita GDP and the Consumption-Output Ratio

119
Appendix 4.1
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’ (exogenous) 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). 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:

\[ c_D = \left( \frac{1}{1+\beta} \right) \left( y_1 + \frac{y_2}{R} \right). \]

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

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

\[ c_D = \left( \frac{1}{1+\beta} \right) \left( \frac{y_2}{R} \right) + \left( \frac{1}{1+\beta} \right) y_1. \]

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., \( c_D = a + by_1 \).

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 households are not debt constrained. But if households 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 $\alpha$ come from, as well as stating the conditions under which either hypothesis may be expected to hold.

Appendix 4.2
The Intertemporal Substitution of Labor Hypothesis

In the model studied in 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:

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

where $\lambda$ and $\beta$ are preference parameters. For these preferences, we can identify three relevant marginal rates of substitution:

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

$$MRS(c_1, c_2) = \frac{1}{\beta c_1}.$$ 

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:

$$\lambda \frac{c_j}{1 - n_j} = z_j \text{ for } j = 1, 2.$$ (30)
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:

$$\frac{1}{\beta} \frac{c_2}{c_1} = R.$$  \hspace{1cm} (31)

These choices will be constrained by an intertemporal budget constraint:

$$c_1 + \frac{c_2}{R} = z_1 n_1 + \frac{z_2 n_2}{R}.$$  \hspace{1cm} (32)

Equations (30), (31) and (32) 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 (30), 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:

$$\frac{c_2}{c_1} = \frac{z_2 (1 - n_2)}{z_1 (1 - n_1)}.$$  \hspace{1cm} (33)

From equation (31), we see that $(c_2/c_1) = R\beta$. This allows us to rewrite the equation above as:

$$R\beta = \frac{z_2 (1 - n_2)}{z_1 (1 - n_1)}.$$  \hspace{1cm} (33)

Equation (33) 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 (33) 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 (33), 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, (33) 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 effect may play a significant role in explaining why employment varies so much across seasons.

### Appendix 4.3

**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 (29), the real interest rate between any two adjacent periods must satisfy:

$$ R^*_1 \ = \ \frac{1}{\beta} \frac{y_2}{y_1}. $$

$$ R^*_2 \ = \ \frac{1}{\beta} \frac{y_3}{y_2}. $$

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-arbitrage 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.
CHAPTER 5
Government Finance

1. Introduction

In this chapter, I utilize the small open economy model developed in the previous chapter to 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.

2. The Government Intertemporal Budget Constraint

As in Chapter 3, I consider an exogenous demand for government purchases. 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, in this dynamic context, the statement ‘an increase in government spending’ is ambiguous. That is, we have 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 \((\tau_1, \tau_2)\).

One financial plan entails period-by-period budget balance; i.e., \(\tau_1 = g_1\) 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 households). 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:

\[
s_G \equiv \tau_1 - g_1. \tag{34}
\]

where \(\tau_1\) represents the government’s current income flow (current tax revenue)
and $g_1$ represents current government purchases of consumer goods and services.

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., $b_G \equiv -s_G \equiv g_1 - \tau_1$. You can think of $b_G$ as representing the government bonds that must be issued to cover the shortfall in tax revenue.

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

$$\tau_2 = g_2 + Rb_G.$$  \hspace{1cm} (35)

By substituting $b_G = g_1 - \tau_1$ into (35), we derive the following expression:

$$g_1 + \frac{g_2}{R} = \tau_1 + \frac{\tau_2}{R}. \hspace{1cm} (36)$$

The expression in (36) 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 (36) 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.

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 $(\tau_1, \tau_2)$ 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 given by:

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

Combining these latter two expressions, we can form the household’s **intertemporal budget constraint**:
\[ c_1 + \frac{c_2}{R} = (y_1 - \tau_1) + \frac{(y_2 - \tau_2)}{R}. \]

Obviously, this is the same as what we derived in Chapter 4, modified to incorporate the household sector’s tax obligation.

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**.

For given interest rate \( R \) and a given after-tax wealth \( W - T \), the solution to the household’s choice problem is depicted in Figure 5.1 (point A).

**FIGURE 5.1**
Individual Choice with Lump-Sum Taxes

While Figure 5.1 looks rather innocuous, it suggests something that will turn out to have important implications for the conduct of government policy. In particular, consider two different **disposable income** profiles; e.g., points B and C in Figure 5.1. Note that while these are two very different looking income profiles, both of them generate the **same** after-tax wealth for the representative household. The implication here is that even a radical change in a household’s
disposable income profile (e.g., from B to C) might have absolutely no effect on consumer demand (if the change leaves disposable wealth unchanged).\(^{44}\)

4. The Ricardian Equivalence Theorem

When an economy is in recession, one often hears calls for the government to reduce taxes to help stimulate consumer demand (by increasing disposable income for the household sector). The tax-cut, it is argued, should be financed by issuing bonds (rather than by reducing government purchases—an act that would further weaken demand at exactly the wrong time). This type of policy is called a **deficit-financed tax cut**.

The intuition for why a bond-financed tax cut should stimulate consumer demand sounds compelling. But does such intuition survive the logic imposed by our theory? The experiment in question holds the pattern of government purchases fixed at some \((g_1, g_2)\). Now, if program spending is left unchanged, then it follows from the government’s intertemporal budget constraint (36) that a cut in current taxes must imply an increase in future taxes. In fact, the future tax increase must be more than the current tax cut, since future taxes will have to accommodate both the principal and interest on the accumulated debt; i.e., \(\Delta \tau_1 < 0 \Rightarrow \Delta \tau_2 = -\Delta \tau_1 R > 0\).

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.

**Exercise 5.1:** Using Figure 5.1, explain how a deficit-financed tax-cut is like moving the disposable income profile from point B to point C. While the deficit-financed tax cut does not affect desired consumer spending, does the same hold true for desired private saving? Explain.

What is going on here? Let me try to put things in a slightly different way. While the current tax cut increases current disposable income of our model households, these households are at the same time forecasting a future tax hike—and hence a reduction in their future disposable income. The increase

\(^{44}\)Of course, this result hinges critically on the assumption that household’s do not face a binding borrowing constraint; see Chapter 4.
in current disposable income and reduction in future disposable income both have the effect of increasing desired private sector saving (see Exercise 5.1). This follows as a straightforward application of the consumption-smoothing motive that we studied extensively in Chapter 4. Since consumer demand remains unchanged, it follows that desired private sector saving rises by the full amount of the tax-cut. In other words, households do not spend their tax-cut; instead they choose to save it. The way they increase their saving is by purchasing the new government bonds that are issued to finance the tax shortfall. When these bonds mature, the household can use these resources to pay for the higher future taxes!

Exercise 5.2: Demonstrate that while a deficit-financed tax cut increases desired private sector saving, it leaves domestic saving (the sum of private and public sector saving) unchanged; i.e., $\Delta s_P = -\Delta s_G = \Delta b_G$. What does this result imply about the effect of an increase in the government budget deficit on the current account balance? Explain.

Exercise 5.3: Using a closed-economy version of this model, demonstrate that a deficit-financed tax-cut will leave both consumption patterns and the equilibrium real rate of interest unchanged. Explain why the increased level of government borrowing has no effect on the interest rate.

The implications of these results are rather startling. In particular, one might provocatively state that deficits do not matter. In particular, note that the government can run its deficit as high as it wants (subject to its intertemporal budget constraint), without having any effect on the economy (and, in particular, on economic welfare). As Robert Barro has stated: Government bonds do not constitute net wealth to the household sector (since these bonds constitute a future tax obligation).

The implications of this particular fiscal policy constitute a part of what has become known as the Ricardian Equivalence Theorem. Loosely speaking, the 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.

The Ricardian Equivalence Theorem asserts the irrelevance of government budget deficits; not the irrelevance of government spending policy.

---

45 The only effect is in the composition of domestic saving between its private and public sector components.

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 include:

1. **No Borrowing Constraints.** In our model, households are free to save and borrow at the market interest rate. One can demonstrate, however, if households face a binding borrowing constraint, then the timing of taxes does affect consumer demand. The question here is mainly an empirical one: do we really believe that significant numbers of households face binding borrowing constraints? If not, then the theorem’s prediction concerning aggregate consumer behavior is likely to be accurate.

2. **Rational Households.** In particular, households must be ‘forward looking’ and understand the government’s intertemporal 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 (generally) 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 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. Robert Barro 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 reasonably good approximations for 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 5.4: Explain why the Ricardian Equivalence Theorem is unlikely to hold in an economy that expects to experience net migration flows (either net immigration or emigration).

Exercise 5.5: Go back to Figure 4.10 and imagine that households face a binding borrowing constraint (so that the desired allocation is at point B). Demonstrate that a government subsidy in the current period (a tax cut financed by a future tax increase) will stimulate current period consumer spending.

5. Government Purchases in an Endowment Economy

It is important to understand that while our model implies that government budget deficits ‘do not matter,’ the same is not true of the government’s spending program. 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.

Once again, keep in mind that we are working here with an endowment economy. What this means is that any increase in government spending will ultimately manifest itself as lower private sector consumption (leaving real GDP unaffected). Once you understand the effect of fiscal policy in this simplified environment, it is relatively easy to extend the model to allow for production and study an additional consequences.

In this section, I consider a policy shock that takes the form of a government announcing a temporary increase in government purchases. To work out the effect of this type of shock, begin with an initial policy \((g_1, g_2) = (0, 0)\). Furthermore, assume that households are initially content with consuming their endowment (recall that the exact position of the initial indifference curve does not matter); i.e., point A in Figure 5.2.
Now, a transitory government spending shock can be modeled as $\Delta g_1 > 0$ and $\Delta g_2 = 0$. There is the question of how the government is to finance the transitory increase in spending. There are two possibilities: [1] tax finance; or [2] bond finance. Let’s assume that the government chooses to tax finance so that $\Delta \tau_1 = \Delta g_1 > 0$.

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](image)

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 trade balance deficit must be matched in the future by a corresponding trade balance surplus (domestic households must export goods and services to the foreign sector to pay back their debt).
Exercise 5.6: Figure 5.2 assumes that the temporary increase in government spending is financed by a corresponding tax. Repeat this exercise, except now assume that the government chooses to bond-finance the increase in spending. Demonstrate that the effects (on consumption and the trade balance) are exactly the same under each finance scenario. Explain.

Exercise 5.7: Suppose that the government announces today that it plans to increase its purchases some time in the future. The government plans to raise taxes only in the future to finance this added expenditure. Explain why such an announcement is likely to depress private consumer spending today (and in the future).

6. Government Policy in a Production Economy

Things start to get a little more interesting once we extend our simple model to allow for both production and distortionary taxation. We can follow the analysis in Chapter 2 and assume that output is produced with a linear production technology \( y_j = z_j n_j \), for \( j = 1, 2 \). Then, using the arguments developed in that chapter, we know that in equilibrium, real wages will be given by \((w^*_1, w^*_2) = (z_1, z_2)\), and profits \((d^*_1, d^*_2) = (0, 0)\).

In a 2-period model, the preferences of households must be modified to include time-dated leisure; i.e., \( u(c_1, l_1, c_2, l_2) \). 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:

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

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 after-tax 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, refer to Chapter 2.

6.1 Ricardian Equivalence Revisited

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 on future employment and output, 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.

Deficits represent future taxes. Taxes tend to be distortionary. Therefore, high deficits today signal potentially high future distortions – and reduced economic activity.

6.2 Government Spending Shocks

When taxes are lump sum, any type of positive government spending shock will 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.

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 \((\tau_1, \tau_2)\) that satisfies the government’s intertemporal budget constraint. However, when taxes are distortionary, Robert Barro has pointed out that it would be optimal for the government to smooth taxes over time.\(^{47}\) 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.

Since taxes are distortionary, the timing of taxes does matter. If government expenditures fluctuate over time, an optimal tax policy entails a relatively constant tax rate, with surpluses and deficits to absorb transitory fluctuations in expenditure.

6.4 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 eco-

nomics 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 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.

7. 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.
Problems

1. 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.

2. Consider an economy populated by two types of individuals, $A$ and $B$. Normalize the total population to unity and let $\theta$ 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 $\tau$ 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?
CHAPTER 6

Capital and Investment

1. Introduction

Let me use a coconut economy as metaphor to help describe what we’ve done so far and where this chapter is designed to take us. By a coconut economy, I mean an economy where all output takes the form of coconuts (I do not mean an economy consisting of entirely of lunatics; although, one might be tempted to model politics in the Middle East on such a premise).

In Chapter 2, we imagined that coconuts were produced with labor effort (together with the existing technology for coconut production). In Chapter 4, we abstracted from labor effort decisions (i.e., by assuming that everyone just worked a fixed number of hours per day) and imagined that coconuts just fell from the sky over time \((y_1, y_2)\). We assumed that coconuts were nonstorable (so that they could not be held as inventory). In other words, coconuts were considered there to be a pure consumption good. You could not inventory coconuts nor use them to help produce future coconuts (e.g., by planting them in the ground).

If output takes the form of a pure consumption good, then there can obviously be no investment (the purchase/construction of new capital goods, including inventory). But even without investment, we were able to speak of saving and the role of financial markets. Even if I cannot inventory a coconut (for future consumption), I might still lend it to someone else in exchange for one of their future coconuts. Here, we see that from an individual perspective, the act of storing a coconut as inventory (investment) and the act of purchasing a bond look very similar: both constitute ways in which to transport consumption from the present to the future. But there is a key distinction when one views these two acts from an aggregate perspective. In particular, the bond purchase does nothing to increase the aggregate supply of coconuts in the future (it simply reallocates a given number of coconuts from debtor to creditor), while the act of investment does. To put things another way: investment constitutes an intertemporal production technology in that it allows an economy to transform current output into future output. This is different from a financial market, which serves simply to rearrange individual claims to a given amount of intertemporal output.

The purpose of this chapter then, is to extend our 2-period model to allow for an intertemporal production technology. Doing so will allow us to examine the role of capital (a stock variable) and investment (a flow variable) in an economy. 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. From Chapter 1, we learned that 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 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. One characteristic of developed economies is the vast quantity of ‘coconuts’ that have been accumulated over time that help produce the material living-standards that we have grown accustomed to.

2. A Model of Inventory Investment

The simplest way to introduce investment into our model is to assume that coconuts are storable across time. To drive the main point home, we can consider a simple Robinson Crusoe economy (a closed economy). As before, Crusoe has preferences \( u(c_1, c_2) \) and an endowment \((y_1, y_2)\). With nonstorable coconuts, Crusoe faced the constraints \( c_1 = y_1, \ c_2 = y_2 \). But now, let \( x \geq 0 \) denote the quantity of coconuts that are held back as inventory; here, \( x \) will represent inventory investment. Assume that inventory depreciates at a constant rate \( 0 \leq \delta \leq 1 \). In this case, the relevant constraints are given by:

\[
\begin{align*}
\ c_1 + x & = y_1; \\
\ c_2 & = y_2 + (1 - \delta)x.
\end{align*}
\]

These two constraints can be combined to form an intertemporal production possibilities frontier (PPF):

\[
\begin{align*}
\ c_2 & = y_2 + (1 - \delta)(y_1 - c_1); \quad (37)
\end{align*}
\]

where \( c_1 \leq y_1 \) (cannot hold negative inventory).

The solution \((c_1^*, c_2^*)\) to Crusoe’s decision problem is depicted as point A in Figure 6.1. Mathematically, this point is characterized by two equations: \( MRS(c_1^*, c_2^*) = (1 - \delta) \) and the PPF \((37)\). Optimal inventory investment can then be calculated as \( x^* = y_1 - c_1^* \).

Notice that the ability to hold inventory is like a technology that allows Crusoe to smooth his consumption over time, even in the absence of any bond market. The (gross) rate of return to inventory investment is given by \((1 - \delta)\).
There is a common perception that the existence of inventory implies an ‘excess supply’ of (or ‘insufficient demand’ for) goods. Such an interpretation appears to have no basis according to the theory presented here. Inventory is simply one way in which agents can store output across time. The ability to store goods allows for (welfare-improving) consumption-smoothing behavior.

3. Capital and Production

3.1 Technology

I now modify our model to allow for the possibility that accumulated coconuts are useful in production. The way we can do this is by following the
lead taken in Appendix 2.1, where we model the production technology as
\[ y = zF(k, n). \]
That is, output (new coconuts) is produced both with labor (\( n \)) and capital (old coconuts) (\( k \)), according to a production function \( F \) with associated technology parameter \( z \).

Because this is a dynamic model, we should in fact write \( y_j = z_j F(k_j, n_j) \) for \( j = 1, 2 \) (current and future production). The current capital stock \( (k_1) \) is assumed to be given exogenously. Let \( x_j \) denote gross investment at date \( j \). If capital depreciates at the constant rate \( \delta \), then the capital stock evolves over time according to:
\[
\begin{align*}
k_2 &= (1 - \delta)k_1 + x_1; \\
k_3 &= (1 - \delta)k_2 + x_2.
\end{align*}
\]

Even with two periods, things can become rather complicated. To avoid these complications, so as to focus on the essentials, let me make a number of simplifying assumptions. First, let’s assume that capital depreciates fully after it is used in production; i.e., set \( \delta = 1 \). In this case, we have \( k_2 = x_1 \) and \( k_3 = x_2 \). In other words, the future capital stock consists entirely of current investment.

Notice that since Crusoe does not value output at date 3, it will be optimal for him to set \( k_3 = 0 \). Of course, given our assumption that capital depreciates fully, this implies that \( x_2 = 0 \) as well. Consequently, we are left with the dynamic equation \( k_2 = x_1 \). Here’s an idea: let’s just redefine things so that \( x \equiv x_1 = k_2 \) and \( k \equiv k_1 \). In other words, \( k \) represents the initial capital stock (exogenous) and \( x \) represents both current investment and the future capital stock (endogenous).

Next, I want to simplify by abstracting from the labor supply decision. In other words, just assume that Crusoe finds it optimal to supply a fixed amount of labor in every period, so that \( n_j^* = 1 \). Gathering our simplifying assumptions together, we can now write \( y_1 = z_1 F(k, 1) \) and \( y_2 = z_2 F(x, 1) \).

I am not quite finished yet. We can simplify notation further by defining the function \( f(x) \equiv F(x, 1) \). Notice that since \( k \) is exogenous, so is the current GDP. Let \( y \equiv z_1 F(k, 1) \) denote this exogenous level of GDP. Future GDP is given by \( y_2 = z_2 f(x) \). Because I want to emphasize the fact that \( z_2 \) represents the expected productivity of (future) capital, let me denote \( z^e \equiv z_2 \). Likewise, let \( y^e \equiv y_2 \). To summarize then, we have an exogenous level of current GDP \( y \), and an expected level of future GDP given by \( y^e = z^e f(x) \). Note that future GDP depends both on future productivity and the current level of investment.

Finally, let’s place some restrictions on the production function \( f \). Here, we make a number of reasonable assumptions. First, assume that capital is essential for production, so that \( f(0) = 0 \). Next, assume that more capital translates into more production; i.e., \( f \) is an increasing function of \( x \). Finally, assume that there are diminishing returns to capital accumulation, so that \( f \) is increasing at a
decreasing rate with \( x \).\(^{48}\)

We are now in a position to define the production possibilities frontier (PPF) for this economy. As in the previous section, begin first with what we know must hold true in each period:

\[
\begin{align*}
    c_1 + x &= y; \\
    c_2 &= zf(x). 
\end{align*}
\]

Combining the first equation with the second yields:

\[
c_2 = zf(y - c_1). \tag{38}
\]

This equation tells us all the \((c_1, c_2)\) combinations that are technologically feasible (and efficient); see Figure 6.2.

**FIGURE 6.2**
Production Possibilities Frontier with Capital Investment

Compare Figure 6.2 with Figure 6.1. They do not look that different. Figure 6.1 has a linear PPF; this linearity just reflects the assumption of a constant depreciation rate. Figure 6.2 has a curved PPF; the curvature here just reflects the assumption of diminishing returns. In both cases, the essential point remains the same: an investment technology allows the economy to transform current

\(^{48}\)Using calculus, we would represent these restrictions by the conditions \( f''(x) < 0 < f'(x) \).
output (via investment expenditure) into future output. The rate of transformation is given by the slope of the PPF. As it turns out, the slope of the PPF here also corresponds to the marginal product of (future) capital, denoted by $MPK(x, z^e)$ in Figure 6.2.

The marginal product of capital is an important concept. Technically, it is given by

$$MPK(x, z^e) = z^e f'(x) > 0$$

and measures the expected (gross) return to current capital spending (investment). Notice that the $MPK$ is an increasing function of $z^e$ (expected productivity) and a decreasing function of $x$ (diminishing returns).

**Exercise 6.1:** Using a diagram similar to Figure 6.2, depict the economy’s PPF for two different levels of expected productivity $z^e$. Explain how an exogenous change in $z^e$ might be interpreted as a ‘news’ shock.

### 3.2 Optimal Investment Expenditure

Let’s summarize. Robinson Crusoe has preferences $u(c_1, c_2)$. He is endowed with coconuts $y$, that he may either consume ($c_1$) or plant in the ground as an investment ($x$); so that $c_1 + x = y$. The investment matures into future coconuts $z^e f(x)$, which are then consumed ($c_2$); so that $c_2 = z^e f(x)$ or $c_2 = z^e (y - c_1)$. What is Crusoe’s best plan of action, given his preferences and given his constraints? The solution to his decision problem $(c_1^*, c_2^*, x^*)$ is depicted by point A in Figure 6.3.

You should be getting pretty good at this by now. Notice that the optimal allocation (point A) is characterized mathematically by two conditions: [1] the slope of the indifference curve is equal to the slope of the PPF ($MRS = MPK$); and [2] the allocation lies on the PPF. The theory developed here can help us understand how an economy’s investment behavior may depend on technology ($y$) and expectations of technology ($z^e$).

**Exercise 6.2:** For preferences given by $MRS = c_2/(\beta c_1)$ and technology given by $MPK = z^e ax^{a-1}$ ($0 < a < 1$), derive the optimal level of investment $x^*$ as a function of the parameters ($\beta, y, z^e, a$). Notice here that $x^*$ does not depend on $z^e$. Explain why. Hint: Refer to the diagram drawn in Exercise 6.1 and note the offsetting substitution and wealth effects that operate on current consumption.

---

49 That is, the MPK tells us the extra amount of future output that is produced by increasing the level of investment by one (small) unit.
3.3 Intertemporal Propagation

One important feature of an economy with productive capital and investment, missing from our earlier formulations, is the concept of **intertemporal** (or **dynamic**) propagation. In plain language, all this refers to is a situation where a purely transitory shock can have long-lasting effects on GDP. The basic idea here is very simple. Consider a transitory productivity shock that increases current GDP. Some of this additional GDP will be consumed, but some of it may be invested as well. The additional investment shows up in the form of a more productive capital in the future. This additional capital adds further to future output, some of which is consumed and some of which is invested. This additional investment adds yet again to future capital...and so on.

To see how this works formally in our simple 2-period model, consider an exogenous shock that increases $y$ (recall that $y = z_1 F(k, 1)$, so we can think of this event as a transitory increase in productivity; i.e., $\Delta z_1 > 0$ and $\Delta z_2 = \Delta z^c = 0$). As usual, begin with some initial allocation; e.g., point A in Figure 6.4. The effect of the shock is to move the PPF ‘to the right’ by an amount $\Delta y$.\(^{50}\) The new indifference curve is likely to be positioned at a point like B.\(^{51}\) This

\(^{50}\)Note that the $y$-intercept on the PPF moves ‘up’ by an amount $z^c f(y + \Delta y) - z^c f(y)$, which is less than $z^c \Delta y$ since $f'' < 0$.

\(^{51}\)This would be consistent, for example, with your answer to Exercise 6.2.
new allocation reflects the usual consumption-smoothing motive; i.e., $\Delta c_1^* > 0$ and $\Delta c_2^* > 0$. However, note that the increase in current consumption is less than the increase in current income; i.e., $\Delta c_1^* < \Delta y$. Since $x^* = y - c_1^*$, this implies an expansion in capital spending; i.e., $\Delta x^* > 0$. The additional capital spending leads to an increase in future GDP since $y_2^* = z^f(x^*)$.

Exercise 6.3: Explain why the MPK at point B in Figure 6.4 is smaller than the MPK at point A. In other words, explain why a positive, but temporary, shock to current GDP leads to a lower rate of return on capital investment.

4. Small Open Economy Analysis

Now, let’s take the Robinson Crusoe economy developed in the previous section and imagine that Crusoe suddenly has access to an international bond market (suppose that he now subscribes to an internet service that allows him to trade online). Let $R$ denote the (gross) real rate of return on a risk-free bond (remember that $R$ is viewed as an exogenous variable for a small open economy).

The existence of a bond market complicates Crusoe’s decision problem since he now has two ways in which to save: bonds and capital. To put things
another way: Crusoe now faces a portfolio allocation problem (in addition to his consumption-saving decision).

Fortunately, there is an easy way to think about decision-making in this scenario. As it turns out, we can proceed in two easy steps. The first step is independent of the second, but the second step depends on the first. The first step involves choosing the level of domestic investment that maximizes Crusoe’s wealth (not utility). The second step takes this maximum wealth as given (together with the prevailing interest rate) and involves choosing the consumption-saving plan that maximizes utility (we’ve already studied this problem before).

4.1 The Investment Choice: Maximizing Wealth

Imagine that Crusoe is contemplating a domestic investment expenditure of \( x_0 \) coconuts. The marginal product of (future) capital is then given by \( MPK(x_0, z^e) \). This is just a number. It tells us how much extra (future) output that can be produced by increasing the level of investment by one more coconut. In other words, it represents the rate of return on investing an extra coconut domestically.

Crusoe has another option, however. He could take this one extra coconut and instead use it to purchase a bond. This bond purchase will return \( R \) future coconuts. In other words, \( R \) represents the rate of return by investing one coconut in the international bond market (of course, the coconut is shipped off as an export to the foreign agent issuing the bond). Keep in mind that Crusoe also has the option of borrowing coconuts at interest rate \( R \).

The question now becomes this: Given that Crusoe has a coconut to invest, should he invest it in domestic capital, or in the foreign bond? The answer is easy: If \( MPK(x_0, z^e) > R \), then invest it domestically. If \( MPK(x_0, z^e) < R \), then export the coconut (in exchange for a bond).

Now, suppose that it is indeed the case that \( MPK(x_0, z^e) > R \). Crusoe would be laughing (all the way to the bank). Why? Well, think about it: Crusoe could borrow (import) one coconut, at the cost of \( R \) future coconuts. If he turns around and invests this imported coconut domestically, it will produce \( MPK(x_0, z^e) \) future coconuts. He would make a (risk-free) profit of \( MPK(x_0, z^e) - R > 0 \) future coconuts. This is called arbitrage. This is great!

The argument just laid out suggests that Crusoe should continue to expand domestic investment (by borrowing coconuts, if necessary) as long as \( MPK(x, z^e) > R \). Keep in mind, however, that the \( MPK \) will decline as \( x \) is increased. This effect just reflects the assumption of diminishing returns to investment (the function \( f \) is strictly concave). Logic suggests that Crusoe will exploit all arbitrage opportunities (and hence maximize wealth) by expanding domestic investment to a point \( x^D \) that satisfies \( MPK(x^D, z^e) = R \).
Now, suppose instead that we are dealing with the case in which $MPK(x_0, z^e) < R$. Once again, there is an arbitrage opportunity. Crusoe could, for example, scale back domestic investment spending by one coconut. Doing so will cost him $MPK(x_0, z^e)$ units of future coconuts. But he could take this coconut and use it to purchase a bond, which yields $R$ future coconuts. This action results in a profit of $R - MPK(x_0, z^e) > 0$ future coconuts. Again, logic suggests that Crusoe should continue to contract domestic investment to a point $x^D$ that satisfies $R = MPK(x^D, z^e)$. At this point, Crusoe will have exploited all arbitrage opportunities (and hence, will be maximizing his wealth).

We can conclude from this discussion that the desired level of domestic investment spending $x^D$ will be determined by the following no-arbitrage-condition:

$$MPK(x^D, z^e) = R. \quad (39)$$

Condition (39) implicitly defines the domestic investment demand function $x^D$. Notice that (39) suggests that investment demand should depend on $R$ and $z^e$ (but not on $y$ or any parameter related to preferences). To emphasize this dependence, we can write $x^D(R, z^e)$.

**FIGURE 6.5**
Wealth Maximization and Investment Demand

Figure 6.5 depicts the solution to the wealth-maximization problem (point A). By choosing investment level $x^D$, Crusoe obtains a maximum wealth $W^*$. Observe that any other investment level (say, point B) results in less wealth.
(e.g., $W_0 < W^*$). Not surprisingly, observe that $MPK > R$ at point B (while $MPK = R$ at point A); so that condition (39) is violated at point B.

The wealth level $W$ is measured in terms of present value (i.e., it is measured in units of current output). Crusoe’s income flow (net of investment expenditure) is given by $(y - x^D, z^e f(x^D))$; the present value of this flow is given by:\footnote{Note that wealth as a function of investment is given by $W(x) = y - x + R^{-1}z^e f(x)$. Maximizing this function with respect to $x$ yields: $W\prime(x) = -x + R^{-1}z^e f'(x)$. Setting $W\prime(x) = 0$, results in condition (39), where $MPK \equiv z^e f'(x)$.}

$$W^* = y - x^D + \frac{z^e f(x^D)}{R}. \quad (40)$$

Once maximum wealth is determined in this way, Crusoe is essentially left with an **intertemporal budget constraint** given by:

$$c_2 = RW^* - Rc_1. \quad (41)$$

Note that $RW^*$ is also a measure of wealth (measured as a future value).

What we have here is a theory of investment demand. We may as well investigate what our theory implies about how domestic investment spending is predicted to react in response to exogenous changes in parameters.

**Exercise 6.4:** For a technology given by $MPK = z^e ax^{a-1}$ ($0 < a < 1$), use condition (39) to derive the investment demand function $x^D$. Demonstrate that $x^D$ is decreasing in $R$ and increasing in $z^e$.

Our first result is somewhat surprising: the demand for investment does **not** depend on $y$. What this implies is that a **transitory** productivity shock ($\Delta z_1 > 0$ and $\Delta z_2 = 0$) has no effect on investment spending. The intuition for this result is as follows. First, note that the return to investment depends on **future** productivity, not **current** productivity. Current productivity would affect investment demand only to the extent that investment must be financed out of current GDP (as would be the case in a closed economy—see, in particular, Figure 6.4). In a small open economy, however, Crusoe is able to borrow (if necessary) to finance any worthwhile domestic investment. The level of current GDP is therefore not important in determining the desired level of investment.\footnote{Of course, this assumes that Crusoe does not face a binding borrowing constraint or other government-legislated capital controls.}

Our second result is that investment demand depends positively on $z^e$. How do we know this? Consider condition (39) for some given $z^e = z_L$; i.e., $MPK(x^D, z_L) = R$. Now imagine increasing $z_L$ to $z_H$. Then, for the same level of investment spending, $MPK(x^D, z_H) > MPK(x^D, z_L) = R$. This implies an arbitrage opportunity. Given the higher expected return to investment, it makes...
sense to expand investment expenditure to the point \( MPK(x^D, z_H) = R \) once more.\(^{54}\)

**Exercise 6.5:** Using a diagram similar to Figure 6.5, plot the PPF for two different levels of expected productivity \( z^e = z_L \) and \( z^e = z_H \) (with \( z_H > z_L \)). Demonstrate that both \( x^D \) and \( W^* \) are larger when \( z^e = z_H \).

Before moving on, I can’t help mention something important here. Note that an exogenous change to \( z^e \) can be interpreted as a ‘news’ shock (i.e., the arrival of new information that leads the private sector to revise its forecast of the future return to capital spending). There is good reason to believe that such shocks are an important feature of real economies, leading to highly volatile movements in investment (which we happen to observe in the data). Since these shocks constitute information, econometricians (looking only at actual productivity) may be led to misinterpret the volatility in investment as being ‘excessive’ (subject, perhaps, to psychological forces, like the ‘animal spirits’ discussed in Chapter 2). Needless to say, such an interpretation is not warranted without further information. In particular, note that in the model developed here, volatile expectations are not inconsistent with rational expectations.

Our final result is that investment demand depends negatively on \( R \). You can demonstrate this quite easily using a diagram like Figure 6.5. The intuition is straightforward too, since \( R \) represents the opportunity cost of investment. In particular, a higher interest rate (say, on bonds) means that it makes sense to divert resources away from investment toward bond purchases until condition (39) is once again satisfied.

### 4.2 The Consumption-Saving Choice: Maximizing Utility

Now that we have characterized the investment choice that maximizes wealth, we can move on to the second step: maximizing utility. You will find this next step very familiar, since it simply deals with choosing a desired consumption-saving plan, subject to a standard intertemporal budget constraint (41)—a problem we studied extensively in Chapter 4. The solution is characterized by an indifference curve tangent to the IBC; see Figure 6.6.

\(^{54}\)Note to instructors: It may be useful here to draw a diagram with the rate of return on the y-axis and the level of investment on the x-axis. Demonstrate a declining MPK schedule together with a constant interest rate. Characterize investment demand by the intersection of these two curves. Show how the MPK schedule shifts upward when \( z^e \) increases.
In Figure 6.6, point A is determined by the optimal investment choice (wealth maximization) and point B is determined by the optimal consumption-saving choice (utility maximization). As in Chapter 4, feel free to place the indifference curve anywhere along the IBC (its exact position will just be a function of preferences). In Figure 6.6, I have placed the indifference curve in a position that implies a trade balance deficit. Let me elaborate on this.

First, let’s review our national income accounting. By definition, \( Y = C + I + G + (X - M) \). As well, we have the definition of domestic saving \( S = Y - C - G \) (assuming that \( G \) takes the form of purchases on consumption). Combining these two definitions, we have:

\[
S = I + (X - M). \tag{42}
\]

What this definition tells us is that in an open economy, domestic saving can be allocated toward two uses; i.e., to finance domestic investment and/or to finance the trade balance (net exports). Using the notation employed in our model, we can write \( s = x + b \), where \( b = (X - M) \) denotes the net amount of foreign bonds purchased. Alternatively, we can write \( b = s - x \). Note that \( b > 0 \) implies
a trade balance surplus (domestic savings are in excess of domestic investment expenditure needs); while \( b < 0 \) implies a trade balance deficit.

Figure 6.6 depicts a scenario where the economy is a net (current) saver; i.e., \( s^D = y - c^D_1 > 0 \) (there is no government in the model). These savings are not sufficient to finance the desired level of domestic investment spending; i.e., \( s^D < x^D \). Consequently, the economy has to import (investment) goods and services to make up the difference (paid for by issuing bonds).

Exercise 6.6: Using a diagram similar to Figure 6.6, depict a point B that is to the left of point A and explain how this implies a current period trade balance surplus.

Exercise 6.7: Again, using the same diagram, depict a point B that implies \( s^D < 0 \).

5. Closed Economy Analysis

In a small open economy analysis, the interest rate \( R \) is viewed as an exogenous variable. As explained in Chapter 4, economists do not generally view prices as being ‘truly’ exogenous; prices are viewed as being determined by supply and demand conditions. It is a simple matter to extend the model (as we did in Chapter 4) to derive a theory of the (real) interest rate. This is done by collecting all of world’s small open economies and aggregating them into a single (world) economy. The world economy is currently closed to interplanetary trade.

The way to think about this is a follows. Consider the situation as depicted in Figure 6.6. This situation cannot describe a general equilibrium. The reason is simple: at the prevailing interest rate \( R \), there is an excess demand for loanable funds; i.e., \( s^D < x^D \) (or, equivalently, \( b^D < 0 \)). In a closed economy, the trade balance must equal zero. We can think of the interest rate as adjusting to ensure that this is the case; i.e.,

\[
R^* = MPK(x^*, z^c) = MRS(c^*_1, c^*_2),
\]

where \( x^* = s^* \), and \( c^*_1 = y - x^* \), \( c^*_2 = z^c f(y - x^*) \). Notice that this corresponds to the situation depicted in Figure 6.3.

6. Interpreting Current Account and Interest Rate Fluctuations

The theory developed in this chapter can be used to interpret country-specific changes in current account positions (small open economy analysis) as well as movements in the real interest rate prevailing in world financial markets (closed economy analysis).
economy analysis). The theory as developed above emphasizes the role played by exogenous changes to productivity (both actual and expected). But the model is easily extended to incorporate exogenous changes in government policies as well.

I admit that for many students, keeping track of all this theory (especially after the course is over) may seem like a daunting prospect. As it turns out, however, much of the theory presented in this chapter (and Chapter 4) can be summarized by way of one simple diagram. I hope that most students will be able to absorb it and keep it in mind (even after the nitty-gritty details underlying its foundations are long forgotten).

The diagram I have in mind consists of two theoretical objects: a desired saving function $s^D$, and a desired investment function, $x^D$. It is useful to think of $s^D$ as the (domestic) supply of loanable funds and $x^D$ as the (domestic) demand for loanable funds.

The theoretical foundations for $s^D$ were developed in Chapter 4. In that chapter, we wrote $s^D(R, y_1, y_2)$ and emphasized the role of savings as a type of ‘shock-absorber.’ We argued there that desired savings is increasing in $y_1$ (a transitory shock to GDP) and decreasing in $y_2$ (an anticipated change in future GDP). The intuition for these results is based entirely on the consumption-smoothing motive. It follows that a persistent (permanent) change in GDP is not likely to change desired saving by very much. This same intuition remains in place if we replace $(y_1, y_2)$ with $(z_1, z_2)$, reflecting our belief that the fundamental parameters are not GDP, but productivity. Thus, let us write $s^D(R, z_1, z_2)$. If desired saving is increasing in the interest rate, then we can draw such a schedule as it is depicted in Figure 6.7.

**Exercise 6.8:** Explain why the domestic supply of loanable funds schedule depicted in Figure 6.7 will ‘shift up’ in response to a transitory increase in productivity, but likely remain unchanged in response to a permanent increase in productivity.

The theoretical foundations for $x^D$ were developed above (don’t tell me you forgot already!). We showed that desired investment spending is declining in the interest rate, does not depend on current productivity, and is increasing in (expected) future productivity.

**Exercise 6.9:** Show how the investment demand schedule shifts in response to ‘good news’ concerning the productivity of future capital.

Now, let us first consider a small open economy that faces an exogenous interest rate $R_0$. Then investment demand is depicted by point A and the supply of domestic saving is depicted by point B. The distance between point A and B $(s^D - x^D)$ measures the trade balance position (it is negative here). Note that these points A and B correspond to the points A and B depicted in Figure 6.6.
Enrique Mendoza documents the following facts concerning small open economies:55

1. The correlation between domestic saving and investment is positive;
2. Domestic saving does not fluctuate as much as investment;
3. The trade balance is countercyclical (tends to move in the opposite direction of GDP over the cycle).

We also know that productivity shocks tend to be highly persistent. Note that, given the persistence observed in productivity shocks, our theory predicts behavior that is consistent with the evidence. Consider, for example, a positive productivity shock that takes the form $\Delta z_1 > \Delta z_2 > 0)$. This shock is likely to shift the saving schedule up by a ‘moderate’ amount (Exercise 6.8) and leads to a shift up in the investment demand schedule (Exercise 6.9). Hence, the correlation between saving and investment is positive (Fact 1). As well, the increase in saving is less than the increase in investment (Fact 2). Finally, note that the trade balance declines (i.e., the deficit position increases); while at the same time, current GDP rises (Fact 3).

---

Thus, according to our theory, a positive persistent shock to productivity leads to an immediate expansion in GDP (Chapter 2). Consumption-smoothing dictates that some of this extra income is saved (Chapter 4). The expected higher return on capital investment expands domestic capital expenditure (Chapter 6). Since domestic saving does not increase by very much, the bulk of the additional investment is financed by foreigners (we increase our imports of investment goods and services, which deteriorates our trade balance position).

Let us now consider a closed economy (e.g., the world economy). In this case, the productivity parameters should be interpreted as worldwide averages. The general equilibrium interest rate is determined by point D in a world without investment; i.e., $s^* = 0$ (Chapter 4). But when investment is possible, the general equilibrium is determined by point C; i.e., $s^* = x^*$.

**Exercise 6.10:** Evidence suggests that the real interest rate is procyclical. Is our theory consistent with the evidence? Explain.

7. 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 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 case of an 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.

Problems

1. Imagine that a small open economy is subject to transitory changes in productivity. Using a diagram similar to Figure 6.7, 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. If a shock hits the economy that sends the trade balance into deficit, does this necessarily imply that economic welfare declines? Explain.
CHAPTER 7

Unemployment

1. Introduction

In this chapter, I take a closer look at the labor market. In Chapter 2, we looked at a model that explained the determination of the level of employment. We saw there how exogenous shocks to productivity might generate employment level fluctuations, as changing incentives induced households to substitute time between market and non-market activities (leisure).

While the simple model in Chapter 2 constitutes a useful starting point, it cannot help us understand at least two very interesting features of the labor market. The first feature is the existence of large gross flows of workers moving into and out of employment at the same time (say, over the course of one month). These gross flows of workers moving into and out of employment are largely offsetting, so that net changes in the level of employment are small by comparison. The second feature is the existence of unemployed workers. As it happens, there are also large gross flows of workers moving into and out of unemployment at all phases of the business cycle.

2. Transitions Into and Out of Employment

2.1 Evidence

Many countries have statistical agencies that perform monthly labor force surveys that measure various aspects of labor market activity. In these surveys, a person is labeled as employed if they report having done any work during the reference period of the survey (e.g., the previous four weeks). 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. The figure also records the average monthly flows of workers into and out of employment.

---

56 For example, the Labour Force Survey in Canada and the Current Population Survey in the United States.
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 affects most people in the same way—as in Chapter 2), 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.

2.2 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:

\[
u(c, l) = \ln(c) + vl,
\]

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:\(^{58}\)

\[
c = wn + a.
\]

Inserting this budget constraint into (43) together with the time constraint \(l = 1 - n\) allows us to rewrite the objective function as:

\[
V(n) = \ln(wn + a) + v(1 - n).
\]

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

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

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

\(^{58}\)Implicit in this budget constraint is the assumption that individuals cannot save or borrow and that there are no insurance markets.
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 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:

$$\ln(w_R + a) = \ln(a) + v. \quad (46)$$

We can solve equation (46) for $w_R$; i.e.,

$$w_R = (e^v - 1)a. \quad (47)$$

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

---

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

---

160
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 (47) 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. Does this make sense? (It should).

Notice that the individual’s labor supply function can also be expressed in terms of their reservation wage; i.e.,
\[
n^* = \begin{cases} 
1 & \text{if } w \geq (e^v - 1)a; \\
0 & \text{if } w < (e^v - 1)a.
\end{cases}
\]

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.
Economic welfare should be evaluated on the basis of broadly-defined consumption; not on how individuals choose to allocate their time across competing activities. There is no a priori reason to believe that high levels of employment necessarily correspond to high levels of social welfare.

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 (47), note that the slope of the reservation wage function is given by $(e^v - 1)a \geq 0$.

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 Zuzanna) 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 Zuzanna 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.

3. Unemployment

3.1 Definition

There are so many misperceptions concerning the way in which unemployment is defined and measured, that it will be worthwhile to take a moment and set things straight.

First, many people (primarily news reporters) rather carelessly refer to the unemployed as jobless (and vice-versa). In fact, jobless individuals are those who are not employed in the market sector. While these individuals are non-employed, they are not technically unemployed. If one insisted on equating the concept of joblessness with unemployment, then we are done as far as theory is concerned (since the model developed in the previous section provides a theory of joblessness).

Another commonly-held view is that an unemployed worker is someone who is not working, but wants to work. This concept makes about as much sense as
defining an employed worker as someone who is working, but **wants** leisure.\textsuperscript{60} Sometimes, this statement is modified to ‘not working, but is willing to work at the market wage.’ The problem with this modification is that it is very difficult to identify what ‘the’ market wage is for any given individual, since people obviously differ along so many dimensions (skill, experience, age, etc.). But in any case, the mere ‘wanting’ of employment (at any wage) is not sufficient to describe the state of unemployment.

The way unemployment is defined and measured is as follows. 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. Note that 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

\textsuperscript{60}In fact, this latter concept makes more sense to me than the former. Throughout the history of mankind, the want of leisure has more commonly been seen as the relevant social problem (rather than a want of work).
ask people whether they ‘want’ leisure but were unable to find leisure. 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.

**FIGURE 7.4**

Average Labor Market Stocks and Monthly Flows  

![Diagram](image)

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 (non-employed 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)—whether or not the going wage is ‘market-clearing’ or not. To explain unemployment, we have to model the reason for why people would willingly choose to allocate scarce time to an activity like job search.

3.2 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’ \( v \)).

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

\[
\begin{align*}
\quad \quad u(c, l) &= c + vl.
\end{align*}
\]

With utility now linear in consumption, this modification implies that wealth will play no role in decision-making.

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 to engage in 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. Remem-
ber that in the event of failure, individuals are free to allocate the remainder of their time \( \theta \) 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 \( w^S_R \) that identifies the type A individual who is just indifferent between work and search; this reservation wage is given by:

\[
\begin{align*}
  w^S_R &= (1 - \theta)(1 - \pi)w' + \pi v.
\end{align*}
\]

Thus, any type A individual with \( w < w^S_R \) 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 \( w^S_R \). Consequently, we can identify a reservation ‘leisure value’ \( v_R \) that just equates the payoff to leisure and search; i.e.,

\[
\begin{align*}
  v_R &= (1 - \theta)(1 - \pi)w' + \pi v_R.
\end{align*}
\]

We can solve the expression above for \( v_R \) as follows:

\[
\begin{align*}
  v_R &= \left[ \frac{(1 - \theta)(1 - \pi)}{1 - (1 - \theta)\pi} \right] w',
\end{align*}
\]

and is plotted in Figure 7.5. Here, any type B individual with \( v < v_R \) will prefer search to 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 is 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 (over the reference period of the survey).

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 7.3: 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.

3.3 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 generally 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 appear to 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.

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.\(^{61}\) By operating a well-designed insurance scheme, it is conceivable that in some circumstances, the government may be able to improve economic welfare.

Exercise 7.4: Many economies have instituted minimum wage laws; i.e., laws that prevent individuals from working at jobs that pay less than some mandated

\(^{61}\)While governments make it difficult for private firms to garnishee human capital, the government itself does not restrict itself in this manner.
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.

4. Summary

Aggregate fluctuations in the aggregate 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) is less than useful.

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, health, 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 their 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 nonpar-
participants 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).
Appendix 7.1: A Dynamic Model of Unemployment and Vacancies

Consider a world with a fixed number of individuals who either work if they have a job or search if they do not (i.e., individuals do not value leisure). The economy is also populated by firms that either produce output (if they have a worker) or expend effort recruiting workers (i.e., if they are vacant). 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 $\theta y$ 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 present value of the expected stream of profits generated by a matched firm is given by:

$$J = \frac{\theta y}{r + s}. \quad (50)$$

**Exercise 7.5:** Explain (i.e., do not simply describe) how the value of a firm depends on the parameters $\theta, y, r$ 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 $\kappa$ 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 $\kappa$ is given by $q(x)J$. If $q(x)J > \kappa$, then it would be worthwhile for more unmatched firms to incur the cost $\kappa$, 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 $\kappa$ and not; i.e.,

$$q(x^*)J = \kappa. \quad (51)$$

The condition above is depicted graphically in Figure 7.6.

**Exercise 7.6:** Using conditions (50) and (51), 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:

$$u_{t+1} = u_t + s(1 - u_t) - p(x^*)u_t;$$

or,

$$u_{t+1} = u_t + s - p(x^*)u_t. \quad (52)$$
Equation (52) is depicted in Figure 7.7 (assuming that $0 < p + s < 1$).

![Equilibrium Unemployment Rate Dynamics and Steady State](image)

**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 (52) to solve for the NRU; i.e., setting $u_{t+1} = u_t = u^*$, we can calculate:

$$u^* = \frac{s}{s + p(x^*)}, \quad (53)$$

**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

Help Wanted Index (Vacancies)  Unemployment Rate
CHAPTER 8

Money

1. Introduction

To this point, I have concentrated on theory designed to explain the determination of \textit{real} economic variables, like the real GDP, the real interest rate, and the real wage without any reference to money. In the models studied so far, people were imagined to exchange goods for goods (like leisure in exchange for output) or goods for claims to goods (like current output in exchange for future output).

In any model of exchange, one is always free to pick an arbitrary good and make it the \textbf{numeraire}. As the name suggests, a numeraire consists an object that serves as a basis for measurement (i.e., as a \textit{unit of account}). In Chapter 2, we implicitly used output as the numeraire good and measured the price of leisure (or labor) in units of output (i.e., the real wage). In Chapter 4, we implicitly used future output as the numeraire good and measured the price of current output in units of future output (the real interest rate).

Note that one is always free to rename a numeraire good; for example, by labelling one unit of output a ‘dollar.’ In fact, the term \textit{dollar} was originally a name given to a specific quantity of gold. But what’s in a name? That which we call ‘gold’ by any other name would shine just as brightly.\footnote{With apologies to Bill Shakespeare.} We could just as well have named the numeraire good ‘Henry’ and it wouldn’t have made any difference. While the monetary objects we call ‘dollars’ today do serve as a unit of account (the way that output does in the models studied earlier), this does not appear to be the primary function of money.

We all have a sense that the primary role of money is to somehow facilitate the exchange process. That is, rather than trade one good for another, it seems easier to first trade one good for money, and then to turn around and trade the money for another good. To put things another way, people seem perfectly willing to trade a good that they value (e.g., leisure) in exchange for an object that they have no intention of ever consuming (money). Evidently, money is not valued for its own sake; but rather, it is valued because it helps them acquire the goods they do desire. In short, money appears to serve as a \textit{medium of exchange}.\footnote{Of course, to the extent that exchange takes time, money must also serve as a \textit{store of value}. This store of value property is not, however, unique to money as any debt instrument or capital good can also serves as a store of value.} But if this is the case, then the models we have studied so far are clearly missing something; in those models, the exchange process works perfectly well without any intermediate monetary exchanges. What are these
models missing?

The purpose of this chapter is to study the theory of money. What makes money essential for exchange? Or, to put things another way: why is there a demand for money?

2. A Model of Private Money

In principle, any type of private debt instrument could serve as means of payment. Suppose you wish to sell your car and that I want to buy it. Suppose further that I offer to pay for the car by issuing a paper note (with my signature on it) promising the noteholder one semester’s worth of economic theory? Perhaps you do not value such a series of lectures. But surely some people do (as evidenced by my overflowing classes). If a sufficient number of people in the economy were confident that: [1] at least someone values my output; and [2] I am willing and able to keep my promises, then why shouldn’t my note (a privately-issued IOU) circulate as a means of payment?64 Such notes could circulate for some time before being redeemed by those individuals who actually value the promised service.

We can formalize the idea above by way of a simple model. Consider an economy consisting of three types of people, labelled A, B and C. Each type 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 $y_3$.

Now, imagine that these people are your kids. How would you instruct them to trade? The answer is obvious: At date 1, instruct B to hand over their endowment to A (who want it); at date 2, instruct C to hand over their endowment to B (who want it); and at date 3, instruct A to hand over their endowment to C (who want it). In this way, everyone gets what they want; such an allocation is efficient (Pareto optimal).

The structure of this economy is interesting because it features what economists call a lack of double coincidence of wants. A lack of double coincidence refers to a situation where no gains to trade exist between any two parties to an exchange. In the context of the earlier example, think of A as representing me and B as representing you. I value your car ($y_1$), but you do not value my future lecture ($y_3$). Left on our own, we would not trade since you have nothing to gain (much to my frustration). The same holds true for any arbitrary pairing of types.

Contrary to popular belief, the lack of double coincidence does not, in itself, imply a role for money. In particular, I have already described how an institution that simply ‘gathered the kids together’ and instructed them how to trade would

---

64 Note that indivisibility is not a problem here as I could issue millions of such notes, with each note entitling the holder to an arbitrarily small fraction of my time.
result in an efficient allocation (without the use of money). Alternatively, any cooperative arrangement could achieve the same thing (people could just get together and agree to do the right thing). Finally, one could also imagine that trade is organized by a financial market in which claims to various time-dated commodities are exchanged (say, at the beginning of date 1). Subsequent to trading, individuals would then just redeem their claims over time. Everyone ends up getting what they want without a medium of exchange; money is not essential.

The existence of money cannot be explained by appealing solely to a lack of double coincidence of wants.

The argument above hints at what it might take to make a monetary instrument essential for trade. In particular, note that for a cooperative arrangement or a financial market to work well, some degree of trust or commitment is necessary. One way or another, people must be either willing or able to keep the promises they make.

One can think of commitment as a technology of sort. A commitment technology is either available, or it is not. In the context of the model above, if a commitment technology is available, then the efficient allocation is achievable even without money (despite the lack of double coincidence of wants). If a commitment technology is not available, then individual rationality dictates that no trade will take place. Such a stark outcome is called autarky.

Let me now modify the model above in the following simple way. First, to prevent a financial market from working ‘too well,’ let’s assume that most people in the economy lack a commitment technology (e.g., types B and C). But to allow at least the possibility of some trade, let’s assume that some people are in possession of a commitment technology (e.g., type A). It is clear that a standard financial market will not work here. This is because type B and C agents can only issue worthless promises.

But there is an alternative trading arrangement that will work. 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,

---

65 Technically, we would have to assume further that each individual values his own good at least a ‘little bit.’ Otherwise, it would be rational to dispose of a good that is not valued at all.
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 will 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 the exchange pattern described above, we can clearly identify a monetary instrument (an IOU issued by $A$). Money is necessary to facilitate exchange in this economy because of a lack of double coincidence and a general lack of commitment. The trading pattern just described is depicted in Figure 8.1.

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. 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

---

66 Would you offer a loan to an anonymous person?
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 is essential in facilitating transactions (without money, there would be no trade). Furthermore, observe that the trading patterns that do occur result in an efficient allocation (i.e., the same allocation that would have occurred if all agents possessed a commitment technology). This last observation provides some justification for why we ignored money in earlier chapters. In particular, to the extent that money/financial markets work well, money can be ignored if we are only interested in understanding the behavior of real allocations.

3. Government Fiat Money

An important (but not dominant) form of money in modern economies takes the form of government fiat; e.g., the small-denomination paper notes commonly used in everyday exchanges. A money is said to be fiat if: [1] it possesses no intrinsic value; [2] it represents no claim against anything of intrinsic value; and [3] costs (virtually) zero to produce. In short, fiat money is intrinsically worthless. Even a similarly-sized sheet of toilet paper has more intrinsic value than that $10 bill in your pocket.

Despite the lack of any intrinsic value, government fiat money obviously circulates. People seem willing to work hard (sacrifice leisure) for these little bits of paper. A visitor from another planet could well be forgiven for questioning how any of this makes sense. A great challenge for monetary theory is to answer the question of how an intrinsically worthless object may nevertheless possess economic value.

**Exercise 8.1:** Explain why a fiat money instrument will not be valued in the A-B-C model studied in the previous section. Hint: consider date 3 and ask what will happen if type C tries to use fiat to purchase the output of type A.

There are two very different perspectives one might take in explaining why government fiat is valued. I call these two views: [1] the legal restrictions hypothesis; and [2] the economic efficiency hypothesis.

The legal restrictions hypothesis asserts that government fiat money is valued primarily because the government ‘forces’ it to have value. Many real-world

---

67 Most people are surprised to learn that only a small fraction of an economy’s money supply actually consists of government money. In Canada (1999), for example, government currency in circulation totalled approximately $32 billion (about $1,000 per person); while demand deposits (electronic money created by private chartered banks) totalled approximately $208 billion (about $6,500 per person).

68 This is in stark contrast to the private money studied in the previous section. The monetary instrument there represented a claim against something of intrinsic value.

69 In fact, the dictionary definition of fiat corresponds to the notion of an arbitrary order,
restrictions can be interpreted as an attempt by the government to bolster the value of its paper. A non-exhaustive list includes: (a) a legal monopoly over the paper money supply; (b) the requirement that private bank money be made demandable for government money (as opposed to other assets); and (c) reserve requirements that force private agencies to hold government money. According to this view, the abolishment of such restrictions would likely drive government money from circulation (being replaced by superior privately-issued debt instruments). Thus, while money may be essential, fiat money is not.

In contrast, the economic efficiency hypothesis argues that there may be circumstances that render a fiat money instrument essential. While this view does not argue one way or the other over who should control the supply of fiat, it seems natural to suppose that an institution like the government takes charge in this matter. Given a ‘well-managed’ supply of fiat, the economic efficiency hypothesis asserts that it will be valued by market participants. This can only be true if fiat money helps to facilitate trades that might not otherwise take place. In the next section, I present a model designed to formalize this idea.

4. A Model of Fiat Money

The model I am about to describe resembles the A-B-C model above, except that I want to consider here an infinite time horizon (i.e., an infinite number of time-dated goods) and an infinite number of different types of people.

Let \( t = 1, 2, 3, \ldots, \infty \) denote time-periods and let \( j = 0, 1, 2, \ldots, \infty \) denote different types of people. Let \( c_j(t) \) denote consumption at date \( t \) by a person of type \( j \).

A type \( j \geq 1 \) person has preferences given by \( u(c_j(t), c_j(t + 1)) \). That is, while there are an infinite number of (time-dated) goods available, each person \( j \geq 1 \) only values two of them; and these two goods are adjacent to each other in time. A type \( j = 0 \) is ‘special’ in that these people are assumed to value good \( c_0(2) \) only.

Note that if we want to, we could interpret individuals of type \( j \geq 1 \) as ‘living’ for two periods only. In this case, it would be natural to interpret \( c_j(t) \) as consumption when ‘young’ and \( c_j(t + 1) \) as consumption when ‘old’; (i.e., current and future consumption, respectively). The special type \( j = 0 \) type people only value consumption when ‘old.’ It would also be natural then to interpret a type as generation. In this case, we could label type \( j = 0 \) as the initial old and type \( j = 1 \) as the initial young, since these are the only two generations ‘alive’ at date \( t = 1 \). As of date \( t = 1 \), the remaining types are viewed as future generations.

Observe that, under this interpretation, it generations appear to overlap with each other. That is, at any point in time \( t \), there are two ‘living’ genera-
tions, consisting of ‘old’ people and ‘young’ people. In the subsequent period, the old ‘die’ and the young become old, with a new generation of young who are ‘born.’ This pattern continues on forever. Because the lives of different generations overlap in this manner, this type of setup (with type interpreted as generation) is sometimes referred to as an overlapping generations (OLG) model. It is important to keep in mind, however, that this interpretation of type as generation is made here only as a matter of convenience; in particular, everything I have to say below will continue to hold if we instead suppose that individuals live forever.  

For the time-being, I consider the simple case of an endowment economy (we will extend the model later to allow for production). To this end, assume that each generation \( j \geq 1 \) has a non-storable endowment \( (y_j(t), y_j(t+1)) = (y,0) \). You can interpret this pattern as people ‘working’ when young (producing \( y \) units of output) and then ‘retiring’ when old (producing 0 units of output). The initial old have an endowment \( y_0(1) = 0 \). The intergenerational pattern of endowments can be represented as follows:

<table>
<thead>
<tr>
<th>Generation (Type)</th>
<th>Time (Time-dated goods)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>↓</td>
<td></td>
</tr>
</tbody>
</table>

The important thing to note about the structure of this model is that it features a lack of double-coincidence of wants (whether or not people live forever). Consider, for example, a pairing of type \( j = 0 \) and \( j = 1 \) individuals (an initial old and an initial young). The initial old value time-dated good \( c(1) \), while the initial young value time-dated goods \( c(1) \) and \( c(2) \). The initial young are endowed with \( c(1) \) and would be willing trade some for good \( c(2) \). The initial old, however, are not endowed with any \( c(2) \). There are no gains to trade here.

The same holds true for an arbitrary pairing of types. Consider, for example, a pairing of type \( j = 1 \) and \( j = 2 \) individuals (an initial young and a member of the next generation). The initial young values goods \( c(1) \) and \( c(2) \); and is endowed with good \( c(1) \). He would be willing to trade some \( c(1) \) for \( c(2) \). The next generation, however, values goods \( c(2) \) and \( c(3) \). In particular, this generation does not value \( c(1) \). Once again, there are no gains to trade here.

**Exercise 8.2:** Consider a pairing of type \( j = 1 \) and \( j = 100 \) individuals (an initial young and a member of generation 100). Demonstrate that there is a lack of double coincidence of wants.

---

70 Again, this is just a warning not to interpret various labels like ‘young’ and ‘old’ too literally. These just happen to be convenient labels. What is important is the assumed structure of preferences.
The last thing to specify is the population structure. Let $N_j$ denote the number of type $j \geq 0$ individuals. Additional structure can be imposed by assuming that $N_{j+1} = nN_j$, where $n > 0$ is a parameter. Again, if type is interpreted as generation and if generations live for two periods only (with the exception of the initial old), then $n$ has the interpretation of being the (gross) population growth rate. To see this, observe that the population at any date $t$ is given by $P_t = N_{t-1} + N_t = (1 + n)N_{t-1}$. The (gross) population growth rate is therefore given by $P_{t+1}/P_t = n$.

4.1 The Golden Rule Allocation

As in the A-B-C model, we can ask what an efficient allocation looks like in this environment. At a minimum, an efficient allocation must satisfy the criterion of Pareto optimality.\footnote{Recall that a Pareto optimal allocation is one in which it would be impossible to make someone better off without making someone else worse off.} One problem with this concept of efficiency is that it does not, in general, help us pin down a unique allocation (although, it obviously helps in dispensing with most allocations).\footnote{Consider an economy consisting of two people (A and B) and 100 coconuts. Then any division of coconuts between A and B that sums to less than 100 is not Pareto optimal. But any division of coconuts summing to 100 is Pareto optimal.} I want to avoid this difficulty by imposing an additional criterion called the Golden Rule property.

In mathematical terms, the Golden Rule property dictates that $c_j(t) = c^y$ for all $j \geq 1$ and $c_j(t+1) = c^o$ for all $j \geq 0$. In other words, let’s restrict attention to allocations in which the young of every generation consume the same amount $c^y$ and where the old of every generation (including the initial old) consume the same amount $c^o$. Note that this does not imply that $c^y = c^o$ (although, such an outcome is possible, if it turns out to be efficient).

The Golden Rule property simplifies things considerably. In particular, it allows us to collapse all $j \geq 1$ generations into a single ‘representative young generation,’ with preferences given by $u(c^y, c^o)$. We can define efficiency then as a feasible allocation $(c^y, c^o)$ that maximizes this utility function. The Golden Rule property also looks after the initial old, whose utility function is given simply by $c^o$.

Now that we know how to rank different allocations $(c^y, c^o)$, we turn to the question of feasibility. This turns out to be easy. The total output (real GDP) at any date $t$ is given by $N_t y$. This output must be divided between the young and old at every date, so that:

$$N_t c^y + N_{t-1} c^o = N_t y.$$  

If we divide both sides of this equation by $N_t$, we may alternatively write this resource constraint as:

$$c^y + \frac{c^o}{n} = y.$$  

\begin{equation} \tag{54} \end{equation}
Notice that this resource constraint looks very much like the lifetime budget constraint studied in Chapter 4, with \( n \) playing the role of the real interest rate (keep in mind, however, that a resource constraint is not a budget constraint).

It turns out that the **Golden Rule allocation** is the **unique solution** to the following decision-problem:

Choose \((c^y, c^o)\) to maximize \(u(c^y, c^o)\) subject to: \(c^y + \frac{c^o}{n} = y\).

The solution to this problem \((c^{y*}, c^{o*})\) is depicted in Figure 8.2 by point B (with point A representing the autarkic allocation).

![FIGURE 8.2](Golden Rule Allocation)

Notice that all individuals are made better off at point B relative to point A. In other words, there are clearly gains to trade here, even if there is a lack of double coincidence of wants. By sacrificing some of their current output (transferring it to the old), the young achieve some degree of consumption-smoothing (as they receive a similar transfer when they become old). Since the initial old get to consume something too, they are also made better off.

One way that society might achieve the Golden Rule allocation is by playing a **gift-giving game**. In the gift-giving game, each young person is transfers a
gift of \( (y - e^{\theta y}) \) units of output to the old. This gift is made \textbf{conditional} on whether the old agent himself made a similar gift in the previous period (with the exception of the initial old, who receive the gift unconditionally). This gift-giving strategy constitutes what game theorists call a \textbf{Nash equilibrium}. All this means is that it is \textbf{individually-rational} for agents to play this gift-giving strategy, conditional on everyone else playing the same strategy. To see this, note that if a young person does not make the transfer, he becomes \textbf{disentitled} of his right to a future transfer (so that he ends up with the autarkic allocation, point A).

What is interesting about the gift-giving equilibrium is that it allows society to achieve the Golden Rule allocation in the absence of monetary exchange (despite the lack of double coincidence of wants). In addition, we do not need to assume that agents possess a commitment technology. But what we do need to assume is the availability of a publicly accessible \textbf{record-keeping technology}. This technology is necessary to record who and who did not make the appropriate gift. Without this record-keeping technology, a young agent could avoid contributing a gift. In the next period, he could then ask for his gift, \textbf{claiming} (falsely) that he made a similar gift in the past. If no record of his past behavior exists, it is impossible to verify such a claim. Since all agents can be expected to behave in a similar manner, gift-giving can no longer be an equilibrium (the only equilibrium, in this case, would be autarky).

What am I telling you all this? Patience, dear reader...patience. The main thing to take away from this for now is that (as pointed out earlier), a lack of double coincidence of wants is not, in itself, sufficient to explain why money is essential. If people could be expected to behave cooperatively, then gift-giving can circumvent the double coincidence problem. I would just show up for work, give my lecture (a gift to you), and not expect any money for it. My compensation would come in the form of future gifts from other members of society (possibly including yourself). If people could not be expected to behave cooperatively, then gift-giving can still achieve the same outcome, provided there was a publicly accessible record-keeping technology (I would only receive my future gift, in this case, if other members of society could verify that I did indeed deliver my lecture). Money is not essential here.

Let’s rule out the idea that all people can be expected to behave cooperatively. Then how seriously can we take this idea of record-keeping? Record-keeping is obviously an important part of everyday life and governs many (non-monetary) exchanges. You cook a meal for a friend one night, and she cooks a meal for you the next. You’re willing to buy the first round of beer for the evening if you’re out with Bob, since Bob has a history of buying his ‘fair’ share of the beer. You’re not willing to buy the first round of beer for David, since David has a history of not reciprocating with similar gifts. It is easy to see how economies consisting of relatively small numbers of people may organize their exchanges without using money at all. Presumably, this is because it is relatively easy to keep track of a small number of individual histories. In small
groups of people, it is difficult to remain anonymous.

Of course, macroeconomics is concerned with the behavior of economies consisting of large numbers of people. In large economies, organizations emerge that do keep track of some personal information. For example, a bank may have a record of your credit history. And the government certainly keeps a record of your tax history. But the idea of one huge publicly accessible record-keeping technology—that somehow manages, without cost or error, to record the trading histories of all individuals—sounds a little far-fetched. In a large economy, it is inevitable that many potential trading opportunities will involve two parties that view each other as anonymous agents. Anonymity implies that you have no record of a person’s history and further, you have no reason to believe that you will ever interact with this person again (which bars the formation of a relationship). Two anonymous agents may trade one good for another, but only if there is a double coincidence of wants. If there is a lack of double coincidence, then no trade is likely to occur.

In the next section, I remove the assumption of a record-keeping technology and assume instead that all agents are anonymous. In the environment we are working with here, the only (non-monetary) equilibrium is, in this case, autarky. I will demonstrate, however, that this same environment allows for a monetary equilibrium that improves upon autarky. In other words, I will show how fiat money is essential for trade.

4.2 Monetary Equilibrium

To develop a monetary equilibrium, I proceed in two steps. First, I describe individual decision-making under the assumption that fiat money has value. In other words, I will begin by assuming that our model people expect money to have possess purchasing power. Second, I will demonstrate how this expectation can turn out to be rational (i.e., consistent with a rational expectations equilibrium). In particular, I will demonstrate how the monetary equilibrium is a self-fulfilling prophesy (see Chapter 2).

To begin then, assume that there is a fixed stock of fiat money $M$ (called ‘dollars’). Fiat money is perfectly divisible, perfectly durable, and cannot be counterfeited. Initially, the stock of fiat is distributed evenly among the initial old, so that they each begin with $M/N_0$ dollars.

At each date $t \geq 1$, there is a competitive spot market where output is exchanged for money. Let $p_t > 0$ denote the dollar spot price of output at date $t$ (i.e., the price-level at date $t$). In the first step of developing a monetary equilibrium, we just view the sequence of prices as exogenous. The assumption that people expect money to possess purchasing power restricts this sequence to satisfy $p_t < \infty$ for all $t \geq 1$.

Exercise 8.3: The price-level is the market value of output measured in units of
money; so that money is the numeraire. Alternatively, one might view output as the numeraire and define \( v_t \) as the market price of money (measured in units of output). Here, \( v_t \) measures the purchasing power of one dollar. Explain why \( v_t = 1/p_t \).

**Hint:** use a soft-drink machine as an example, where soft-drinks sell for \( p_t = 0.50 \) dollars. How many soft-drinks must the machine sell to purchase one dollar?

Now, let’s consider the decision-problem facing a young agent at any date \( t \). This person may, if he so chooses, sell some of his output \((y - c^y)\) for money. In doing so, he would collect \( m_t \) dollars, where:

\[
m_t = p_t (y - c^y). \quad (55)
\]

He could then turn around and, in the next period, use his money to purchase output \( c^o \) at the price \( p_{t+1} \). In this next period, his budget constraint is given by:

\[
p_{t+1} c^o = m_t. \quad (56)
\]

Combining equations (55) and (56) implies the following **intertemporal budget constraint**:

\[
c^y + \left( \frac{p_{t+1}}{p_t} \right) c^o = y. \quad (57)
\]

Equation (57) describes a young agent’s **ability to substitute** current consumption \((c^y)\) for future consumption \((c^o)\). The ratio \((p_{t+1}/p_t)\) is an exchange rate (i.e., a relative price). In particular, it constitutes the price of \( c^o \) measured in units of \( c^y \). It also corresponds to the definition of the (gross) **inflation rate** (the rate of change in the price-level over time).

**Exercise 8.4:** Let \( v_t = 1/p_t \) denote the value of money (see Exercise 8.3). Then the (gross) real rate of return on money is given by \((v_{t+1}/v_t)\). Explain how the real return on money is related to the inflation rate.

Let’s stop and think for a moment. In our model, fiat money does not pay interest. Nevertheless, it would be wrong to conclude that money does not generate a return. The **real** return on your money depends on its future purchasing power; i.e., \( v_{t+1} \). The **real** cost of acquiring one dollar today is measured by the current purchasing power that is sacrificed in doing so; i.e., \( v_t \). The (gross) real rate of return on non-interest-bearing money is therefore given by \((v_{t+1}/v_t)\); which is simply the inverse of the (gross) inflation rate.

**Inflation** refers to a price-level that rises over time; i.e., \( p_{t+1}/p_t > 1 \). During an inflation then, we see that the value of money (i.e., its purchasing power) declines over time; i.e., \( v_{t+1}/v_t < 1 \). **Deflation** refers to a price-level that falls over time; i.e., \( p_{t+1}/p_t < 1 \). During a deflation then, we see that the value of
money increases over time; i.e., \( v_{t+1}/v_t > 1 \).

Non-interest-bearing money earns a real rate of return that is related inversely to the inflation rate.

Let \( \Pi_{t+1} \) denote the (gross) rate of inflation; i.e., \( \Pi_{t+1} \equiv p_{t+1}/p_t \). In what follows, I focus on ‘stationary’ equilibria, so that \( \Pi_{t+1} = \Pi \) (i.e., the inflation rate is constant). Then we can rewrite the intertemporal budget constraint (57) as: \( c^o = \Pi^{-1} y - \Pi^{-1} c^y \) and depict it graphically; i.e., see Figure 8.3. The decision-problem facing a representative young person can be stated formally as:

Choose \((c^y, c^o)\) to maximize \( u(c^y, c^o) \) subject to: \( c^y + \Pi c^o = y \).

The solution to this problem is depicted in Figure 8.3 by point A.

If you haven’t noticed by now, you should note the resemblance between this analysis and the one in Chapter 4. In that chapter, we considered a person with an endowment \((y_1, y_2)\) facing an interest rate \( R \) and presented with the problem of choosing the best \((c_1, c_2)\). In this chapter, we have a person with an endowment \((y, 0)\) facing an inflation rate \( \Pi \) and presented with the problem of choosing the best \((c^y, c^o)\). The solution in both cases generates a desired saving function. In Chapter 4, saving took the form of buying or selling real bonds (there was no role for money). In this chapter, savings must take the form of money purchases (because of the lack of double coincidence and anonymity of agents).

Let \( q_t \equiv m_t/p_t \) (i.e., the purchasing power of \( m_t \) dollars (real money balances). Then the desired saving function associated with point A in Figure 8.3 is given by \( q^D = y - c^y \) (the demand for real money balances). Note that this money demand function is predicted to depend on both \( y \) and \( \Pi \).

When we want to emphasize this dependence, we can write \( q^D(y, \Pi) \).

Exercise 8.5: Assuming that \((c^y, c^o)\) are normal goods, how is the demand for real money balances \( q^D \) predicted to change with an exogenous increase in \( y \)? How

\[73\text{The net rate of inflation is defined as } \pi = \Pi - 1; \text{ or} \]

\[ \pi = \left( \frac{p_{t+1} - p_t}{p_t} \right). \]
does the demand for money depend on the inflation rate? Explain (provide intuition).

Exercise 8.6: Suppose that preferences are such that \( MRS = c^o/\beta c^y \). Show that for these preferences, \( q^D = \left( \frac{\beta}{1+\beta} \right) y \). Hint: observe (from Figure 8.3) that \( MRS(c^o, c^y) = \Pi^{-1} \). Explain why the demand for money does not depend on inflation here (a diagram may prove useful).

Exercise 8.7: Explain whether there is any logic in labelling \( c^o \) a ‘cash’ good and \( c^y \) a ‘non-cash’ good.

This essentially concludes step one. To reiterate, we began by supposing that young agents believe that money has future purchasing power. In other words, agents believe that money has some positive rate of return; i.e., that \( \Pi < \infty \) (you can see from Figure 8.3 that if \( \Pi = \infty \), then \( q^D = 0 \)). Given this belief, young agents find it optimal to sacrifice some of their current endowment in exchange for intrinsically worthless bits of paper. This paper is purchased from the old agents (who, aside from the initial old, acquired it in an earlier transaction when they were young).

Observe that handing over output for paper has the smell of a gift. You go to some old bugger’s home and wash his windows (sacrificing leisure) in exchange
for an intrinsically worthless token. Why? Because you believe that this token will have future purchasing power. Can such a belief be rational? To answer this question, we move to step two.

Properly speaking, the inflation rate above is an expected inflation rate. I should have used the notation $\Pi^e$ to make this explicit (sorry). In this case, the money demand function should be written as $q^D(y, \Pi^e)$. Step two involves defining what we mean by a monetary equilibrium. A monetary equilibrium requires the following restrictions to hold:

1. Given some $\Pi^e$, agents choose $q^D(y, \Pi^e)$ optimally (this was described in step one);
2. Given the behavior of agents $q^D$, markets clear at every date $t \geq 1$; and
3. Expectations are consistent with reality; i.e., $\Pi^e = \Pi^* < \infty$ (i.e., expectations are rational).

Condition 1 is summarized by a $q^D(y, \Pi^e)$ satisfying point A in Figure 8.3. Condition 2 requires that markets clear at each date. Recall that the market here is a spot market in which money trades for output. The aggregate demand for real money balances is given by $N_t q^D$. The aggregate supply of real money balances is given by $M/p_t$. An equilibrium price-level $p^*_t$ is one that clears the market; i.e.,

$$\frac{M}{p^*_t} = N_t q^D(y, \Pi^e).$$

(58)

Since this condition must hold at every date $t \geq 1$, the following must also be true:

$$\frac{M}{p^*_{t+1}} = N_{t+1} q^D(y, \Pi^e).$$

Dividing the former by the latter, we derive:

$$\frac{p^*_{t+1} M}{p^*_t M} = \frac{N_t q^D(y, \Pi^e)}{N_{t+1} q^D(y, \Pi^e)};$$

which simplifies to:

$$\Pi^* = \frac{1}{n}. \quad (59)$$

Finally, condition 3 asserts that expectations must be rational, so that $\Pi^e = \Pi^* = n^{-1}$. We can plug this into equation (58) to derive an expression for the equilibrium price-level at each date; i.e.,

$$p^*_t = \frac{M}{N_t q^D(y, n^{-1})}. \quad (60)$$
What we have just demonstrated here is that there exists a monetary equilibrium, with a sequence of price-levels given by equation (60). If individuals expect an inflation rate equal to \( \Pi^e = n^{-1} \), then their optimal behavior will in aggregate result in an equilibrium inflation rate equal to \( \Pi^* = n^{-1} \). Since \( n > 0 \), it follows that \( \Pi^* < \infty \) (i.e., it is rational to expect money to have future purchasing power).

What is remarkable about this monetary equilibrium is that the equilibrium allocation corresponds to the Golden Rule allocation. To verify this fact, note that the equilibrium budget constraint is given by: \( c^o = ny - nc^y \), which corresponds exactly to the resource constraint in Figure 8.2. Since \( MRS(c^y, c^o) = n \) in equilibrium and for the Golden Rule, the allocations must be the same.

4.3 Non-monetary Equilibrium

Since fiat money is, by definition, an unbacked security (an intrinsically worthless claim), it’s value can arise only to the extent that it facilitates exchange. To facilitate transactions, people have to believe that money will retain value in the future. In short, since the value of fiat money cannot be based on collateral, it’s value must be based on societal beliefs. If everyone believes that fiat money has value, then fiat money will have value. The monetary equilibrium then, has the flavor of a self-fulfilling prophesy.

But self-fulfilling prophesies depend on an initial (exogenous) belief structure. As it turns out, our model (as with any model of fiat money) has another rational expectations equilibrium where money is not valued. To construct this non-monetary equilibrium, assume that (for some unexplained reason) people do not believe that paper will be accepted in future exchanges. That is, for some given \( p_t \), suppose that society believes \( p_{t+1} = \infty \), so that \( \Pi^e = \infty \). In this case, individual rationality implies \( q^D = 0 \). From the market-clearing condition (58), we see that this must imply \( P^*_t = \infty \) or \( v^*_t = 0 \) (the purchasing power of money falls to zero, since no one in their right mind would want to accept it in any exchange). Thus, if everyone believes that fiat money is worthless, then fiat money will be worthless (this will be a rational expectation). The non-monetary equilibrium, in this case, is autarky.

5. Money Neutrality

One of the oldest debates in macroeconomics concerns the question of whether money is neutral. Money is said to be neutral if an unanticipated one-time increase in the level of the money supply has no effect on real economic variables. A distinction is often made between short-run and long-run money neutrality. In particular, many economists believe that while money is neutral
in the ‘long-run,’ it is not necessarily neutral in the ‘short-run.’ We can address
the issue of money neutrality by appealing to the theory developed above.

Long-Run Neutrality

To begin, imagine that the economy is initially in some monetary equilib-
rium; i.e., as described above. Let’s assume a constant population \( n = 1 \), so
that \( \Pi^e = \Pi^* = 1 \) (price-level stability). Let \( M_L \) denote the stock of fiat money.
From equation (60), it follows that the equilibrium price-level is given by:

\[
p^*_t = \frac{M_L}{Nq^D(y, 1)} \equiv p_L.
\]

The representative young agent enjoys the allocation \( c^y = y - q^D(y, 1) \), \( c^{*y} = q^D(y, 1) \). The initial old enjoy consumption \( c^{*o} = q^D(y, 1) \).

Evaluating the long-run consequences of a one-time increase in the money
supply is easy. Because the new money supply \( M_H > M_L \) remains constant over
time, inflation expectations remain unchanged at \( \Pi^e = 1 \). Consequently, the
demand for real money balances remains unchanged at \( q^D(y, 1) \). From equation
(60), it follows that the new price level is given by:

\[
p^*_t = \frac{M_H}{Nq^D(y, 1)} \equiv p_H,
\]

where \( p_H > p_L \).

The long-run impact of an increase in the money supply is to leave all real
variables unchanged; hence, money is neutral in the long-run. The only long-run
effect of this shock is to increase the price-level (and all nominal variables) in
proportion to the increase in money supply. The logic here is compelling: How
can the choice of how many zeros to place on government paper money have
any real long-run effects? I am aware of no compelling empirical evidence that
overturns this prediction.

Short-Run Neutrality

Investigating the short-run consequences of any shock is not as easy, since in
general, there may be transition dynamics as the economy moves from one
long-run equilibrium to another. Still, the model here is fairly simple, so we
should be able to figure some things out.

Let me begin by describing the initial monetary equilibrium in greater detail.
Because of the potential for transition dynamics, we have to be more careful with
notation. Recall that \( c_j(t) \) represents the consumption of a type (generation) \( j \)
person at date \( t \). For a constant population, the resource constraint at any

192
The budget constraint facing a young person at date $t$ is given by:

$$c_j(t) + \Pi_{t+1}c_j(t+1) = y.$$  

If $\Pi_{t+1} = 1$ (as it is in the initial equilibrium), then optimizing behavior on the part of the young implies $MRS(c_j(t), c_j(t+1)) = 1$ and $c_j(t) + c_j(t+1) = y$. This implies a demand for real money balances $q_j(t) = y - c_j(t)$. The old of generation $j-1$ get to consume $c_{j-1}(t) = q_j(t)$, which implies $c_{j-1}(t) = y - c_j(t)$. Of course, this is just the resource constraint. Hence, the equilibrium allocation $c_{j-1}(t), c_j(t), c_j(t+1)$ must satisfy both the resource constraint and the budget constraint simultaneously. When $\Pi_{t+1} = 1$, the only way this can happen is if $c_j(t) = c^o$ and $c_{j-1}(t) = c_j(t+1) = c^o$. This is depicted in Figure 8.4.

FIGURE 8.4
Stationary Monetary Equilibrium
(n = 1)

The thought experiment I have in mind is as follows. First, imagine that the economy starts off in an initial long-run equilibrium with money supply $M_L$ and price-level $p_L$ (e.g., as depicted by point A in Figure 8.4). Then, at some arbitrary date $t$, the government ‘surprises’ agents by printing new
money \((M_H - M_L)\). Here, we have to be careful to describe exactly how the government injects the new money into the economy. Let me assume here that the new money is distributed to the old at date \(t\) in as a lump-sum transfer \(A_t \equiv (M_H - M_L)\) (with each old agent receiving \(A_t/N\) dollars). Keep in mind that this happens only once (at date \(t\)) and that everyone is assumed to understand this. For all dates beyond \(t\), the money supply remains constant at \(M_H\).

For this one-time money shock to be neutral, there can be no effect on real variables. In terms of Figure 8.4, this requires that the equilibrium allocation at point \(A\) remain unchanged. For this to be true, the money shock must have no effect on the actual/expected future inflation rate \(\Pi_{t+1} = p_{t+1}/p_t\). Note that this does not imply that the shock must have no effect on the actual current inflation rate \(\Pi_t = p_t/p_{t-1}\). That is, it is possible (and indeed, likely) that this unexpected shock renders \(\Pi_t = \Pi_{t+1}\).

Now, since the young at date \(t\) do not receive the surprise money transfer, their choice problem remains unchanged and we can summarize their optimal behavior by \(q^D(y, \Pi_{t+1})\). The old at date \(t\) do receive the money transfer \(A_t\), but their choice problem at this stage is trivial:

\[
c_{j-1}(t) = \frac{M_L + A_t}{Np_t} = \frac{M_H}{Np_t}. \tag{61}
\]

From the market-clearing condition, we know that:

\[
p_t = \frac{M_H}{Nq^D(y, \Pi_{t+1})}. \tag{62}
\]

This condition must hold at date \(t+1\) as well, so that:

\[
p_{t+1} = \frac{M_H}{Nq^D(y, \Pi_{t+2})}. \tag{63}
\]

Combining these two equations implies:

\[
\frac{p_{t+1}}{p_t} = \Pi_{t+1} = \frac{q^D(y, \Pi_{t+1})}{q^D(y, \Pi_{t+2})}
\]

In the expression above, \(q^D(y, \Pi_{t+2})\) refers to the optimal behavior of the future young generation \(j+1\). At date \(t+1\), the current young will be old and will, at that date, be in full possession of the money supply \(M_H\). Because the current money shock is a one-time event, the money supply is expected at date \(t+1\) to remain constant forever. It seems reasonable to suppose then that, as of date \(t+1\), the economy is its new long-run equilibrium; i.e., \(\Pi_{t+2} = 1\) (with \(p_{t+1} = p_{t+2} = p_H\)). Using the previous equation, this then implies:

\[
\Pi_{t+1} = \frac{q^D(y, \Pi_{t+1})}{q^D(y, 1)}. \tag{64}
\]

Clearly, \(\Pi_{t+1} = 1\) is a solution to equation (64).\(^4\) This obviously implies

\(^4\)If \(q^D\) is weakly decreasing in \(\Pi\), then it is easy to see that this solution is unique.
that $p_t = p_H$. In other words, the equilibrium price-level jumps up immediately to its new long-run level (with no change in inflation expectations). While the old get some money, they compete among themselves to acquire output from the young, which drives the price-level up to a point that leaves their real money balances unchanged. Since the price-level rises in proportion to the money supply at date $t$, it follows from equation (61) that there is no change in the consumption level for the initial old. It is easy to see that the allocation of the young remains unchanged too. Money is neutral in this experiment.

**Short-Run Non-Neutrality**

There is an almost universal belief among economists that for money to non-neutral, some degree of ‘price-stickiness’ is required. In the present context, price-stickiness refers to an assumption that the price-level $p_t$ cannot respond immediately to the date $j$ money shock. If this is true, then one can see clearly from equation (61) how a one-time money shock can have a real effect. In particular, with a fixed price-level, the money transfer considered above increases the purchasing power of the initial old.

In what follows, I develop an example of non-neutrality that does not depend on sticky prices. To this end, let us reconsider the experiment performed in the previous section, but with one minor modification. In particular, assume that the new money is now transferred to the initial young, instead of the initial old. Because the initial young are now the recipients of the transfer, they are affected directly by the money shock. This means that we’ll have to reformulate their decision problem. To this end, let $a_t \equiv A_t/(N_p)$ denote the purchasing power of the monetary transfer accruing to each young agent at date $t$ (recall once again that this transfer occurs only once). Then the decision problem for a date $t$ young agent is given by:

Choose $(c_j(t), c_j(t+1))$ to maximize $u(c_j(t), c_j(t+1))$
subject to: $c_j(t+1) + \Pi_{t+1} c_j(t+1) = y + a_t$.

The solution to this problem can be summarized by a demand for real money balances $q^D(y + a_t, \Pi_{t+1})$.

**Exercise 8.8:** Using a diagram, demonstrate how $q^D$ is an increasing function of $a_t$.

Explain the economic intuition underlying this result (assume that time-dated consumption is a normal good).

Since the date $j$ old agents are not recipients of the transfer, equation (61) must now read as follows:

$$c_{j-1}(t) = \frac{M_L}{p_t}.$$  \hspace{1cm} (65)
Notice here that any change in the price-level will have an effect on consumption (this would obviously not be the case if we just assumed a sticky price-level!).

The first thing we can demonstrate (without getting too technical) is that this money shock cannot be neutral. How do I know this? Consider the following argument.

1. For the money shock to be neutral, the equilibrium allocation must remain unchanged; i.e., it must remain at point A in Figure 8.5;
2. Since money is valued (i.e., $p_t < \infty$), it follows that $a_t = A_t/(Np_t) > 0$;
3. The first two points above imply that money neutrality requires $\Pi_{t+1} > 1$, as this is the only way the equilibrium budget constraint can pass through point A; i.e., see Figure 8.5. In other words, allocation A must be both resource and budget feasible.
4. But if the budget constraint passes through point A, then young agents will not find it optimal to choose this allocation (they will choose point B instead).

FIGURE 8.5
Proof of Money Non-Neutrality
The argument above demonstrates that the Golden rule allocation cannot be a part of the equilibrium allocation in the period of the shock. However, it does not demonstrate what the date $t$ equilibrium looks like. Let’s see if we can figure this out.

Consider this economy at date $t+1$. At the beginning of this date, the old (who were young at date $t$) are in possession of the money supply $M_H$. From this date on, the money supply remains fixed (so that $a_{t+1} = 0$). It seems reasonable to suppose that from this date on, the economy reverts back to the Golden rule allocation (with a higher price-level $p_H$). Thus, $p_{t+1} = p_H$ and $\Pi_{t+2} = 1$.

The consumption for the old at date $t+1$ is given by:

$$c_j(t+1) = \frac{M_H}{N p_{t+1}}$$

But since $p_{t+1} = p_H$, this is just given by:

$$c_j(t+1) = \frac{M_H}{N p_H} = q^D(y + a_{t+1}, \Pi_{t+2}) = q^D(y, 1) = c^o.$$  \hspace{1cm} (66)

Now, let’s consider date $t$. We know that the equilibrium value for $a_t$ must be positive. From exercise 8.8, we know that if $\Pi_{t+1} = 1$, then the consumption profile desired by the young at date $t$ must feature $c_j(t) > c^o$ and $c_j(t+1) > c^o$. This last inequality is inconsistent with equation (66). We can conclude from this that $\Pi_{t+1} \neq 1$. In particular, to ensure that $c_j(t+1) = c^o$, we need $\Pi_{t+1} > 1$. Evidently, the date $t$ equilibrium must satisfy an allocation like point B in Figure 8.6.

Now, since $\Pi_{t+1} = p_{t+1}/p_t = p_H/p_t > 1$, it follows that $p_t < p_H$. Point C in Figure 8.6 depicts the consumption of the initial old. Observe here that $c_{j-1}(t) < c^o$. We know then that the following is true:

$$c^o = \frac{M_L}{N p_L} > \frac{M_L}{N p_t} = c_{j-1}(t).$$

This can only be true if $p_t > p_L$. Thus, we have established the following:

$$p_L < p_t < p_H.$$  

In other words, the price-level jumps up in response to the money shock, but not to its higher long-run value. The surprise money injection results in a surprise inflation; i.e., $\Pi_t = p_t/p_L > \Pi_t^* = 1$. The shock also produces an anticipated inflation; i.e., $\Pi_{t+1} = p_H/p_t > 1$. 

197
Alright, so maybe this isn’t the easiest thing in the world. (Now you know why I started out with a simple static model in Chapter 2!). Let me lay down the basic intuition. The economy starts out with a money supply that is distributed unevenly throughout the population (the old have lots of it and the young have none). The government then gives some new money to the young, thereby adding to the total money supply. The increase in the supply of money lowers its value (i.e., increases the price-level). This leads to a reduction in the purchasing power for those who have money, but did not receive a transfer. In the model, this loss of purchasing power falls on the old and is transferred to the young.

The important points to take away from this are the following:

1. The assumption of sticky-prices is not necessary to explain the (short-run) non-neutrality of money; and

2. The effect of any given monetary injection can depend critically on how the new money is injected.
5. Summary

Money is an object that circulates widely as a medium of exchange. To the extent that exchange takes time, money must also serve as a store of value. Money can also serve as a unit of account, but this does not appear to an essential property of money. In particular, in the model of fiat money studied above, we were free to choose either money or output as the numeraire without affecting the essential properties of the monetary equilibrium (it did not matter whether we stated prices in terms of $p_t$ or $v_t$, since $p_t = 1/v_t$).

What makes money essential in facilitating exchanges? A necessary—but not sufficient—condition is a lack of double coincidence of wants. A further condition requires that at least some trading opportunities involve parties who are anonymous to each other. Anonymity is equivalent to the lack of a societal memory bank that records (and makes available) the private trading histories of individuals. Money can be interpreted as a substitute for this missing record-keeping technology. Absent counterfeit and theft, my money holdings serve as a publicly-observable record of some past sacrifice on my part (i.e., a gift of goods or services made to some member of society). As a reward for this gift, society is willing to reciprocate in a similar manner (by accepting my money for intrinsically valuable goods and services). In this way, monetary exchange appears to replicate an institution based on gift-exchange. Money is essential because agents cannot be trusted to reveal their true individual histories. In short, the essential role of money is to serve as a form of societal memory.\footnote{See: Kocherlakota, Narayana (1998). “The Technological Role of Fiat Money,” Federal Reserve Bank of Minneapolis Quarterly Review, 22(3): 2–10.}
Problems

1. The last several years have witnessed rapid technological development in IT (information technologies), leading some to proclaim the coming of a cashless economy (i.e., an economy in which $q^D$ approaches zero). Use the analysis in this chapter to explain how IT development may reduce the demand for cash.

2. You are planning an extramarital affair (hypothetically speaking, of course). Do you pay for the planned getaway with cash or credit? What does this example suggest about the demand for anonymity (and hence the demand for cash, even in a world where perfect record-keeping is available)?

3. “The quintessential crime of the information age is identity theft, the malicious use of personal identifying data.” Explain how the monetary equilibrium described above can be interpreted as arising in a world where individual identities can be stolen with impunity. Hint: think of the problem identity theft poses in a gift-giving economy.

4. Is deflation ‘bad’ for the economy?

CHAPTER 9

Inflation

1. Introduction

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 monopoly control over the printing press (the supply of small denomination fiat 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.

2. Government Spending and Inflation

Let’s begin with the simple OLG model developed in the previous chapter. A representative young agent has preferences \( u(c_y, c_o) \) and there is an initial old generation that cares only for \( c^o \). The young have a non-storable endowment \( y > 0 \) and the initial old are endowed with fiat money. The number of young agents at date \( t \) is given by \( N_t \) and grows exogenously according to \( N_{t+1} = nN_t \).

Let’s now add to this economy a government that requires \( g \) units of output per young person (so that total real government purchases equals \( gN_t \) units of output, for some \( 0 \leq g < y \)). To focus on money creation as a revenue device, I 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 \( \mu \) 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:

\[
p_t gN_t = \left[ 1 - \frac{1}{\mu} \right] M_t;
\]

or, dividing through by \( p_t N_t \):

\[
g = \left[ 1 - \frac{1}{\mu} \right] \frac{M_t}{N_t p_t}.
\]

For a given (expected) inflation rate \( \Pi^e \), a young agent’s decision problem remains the same as described earlier; i.e., the solution is given by a demand
for real money balances $q^D(y, \Pi^e)$, described by Figure 8.3. The assumption of rational expectations implies that $\Pi^e = \Pi$ (the actual inflation rate).

The final element of the model is the market-clearing condition, which requires the supply of real money balances to equal the demand for real money balances at every date $t \geq 1$:

$$\frac{M_t}{p_t} = N_t q^D(y, \Pi). \quad (68)$$

To construct the monetary equilibrium, begin by noting that since (68) must hold at every date, the following must be true:

$$\frac{M_{t+1}}{p_{t+1}} = N_{t+1} q^D(y, \Pi). \quad (69)$$

Now divide the former equation by the latter:

$$\frac{M_t}{M_{t+1}} \frac{p_{t+1}}{p_t} = \frac{N_t}{N_{t+1}} q^D(y, \Pi).$$

A simple rearrangement of terms and use of definitions gives us an expression for the equilibrium inflation rate (as a function of $\mu$, which is yet to be determined):

$$\Pi^* = \frac{\mu}{n}. \quad (70)$$

Combining equation (70) with the market-clearing condition (68) implies $q^D = M_t / (N_t p_t)$. Plug this into the government budget constraint (67), so that we are left with:

$$g = \left[ 1 - \frac{1}{\mu^*} \right] q^D \left( y, \frac{\mu^*}{n} \right). \quad (71)$$

This equation implicitly defines the equilibrium growth rate of the money supply $\mu^*$ that is consistent with: [1] individual optimization; [2] market-clearing; [3] rational expectations; and [4] government budget balance (given an exogenous $g$).

Let’s look at condition (71) more closely. The left-hand-side represents the value of government spending (per young person). The right-hand-side represents the purchasing power of new money. Note that while the government does not levy any direct taxes in this economy, it nevertheless is able to acquire resources by printing new money. This method of acquiring resources is called seigniorage and is available to government solely because of two facts: [1] people demand money (Chapter 8 explains why); and [2] the government maintains monopoly control over the money supply. Regardless of the name we attach to this power, it should be clear that seigniorage is a tax; this tax is commonly referred to as an inflation-tax.
What effect does this monetary/fiscal policy have on economic behavior and the welfare of individuals? We can answer this question with the help of a diagram. Suppose initially that \( g = 0 \). Then the monetary equilibrium corresponds to the one in the previous chapter. In particular, note that \( g = 0 \) implies \( \mu^* = 1 \), so that \( \Pi^* = 1/n \). In this case, the equilibrium budget constraint lies on top of the resource constraint and the equilibrium allocation corresponds to the Golden Rule allocation with \( MRS(c^o, c^y) = n \). This is given by point A in Figure 9.1.

Suppose now that \( g > 0 \). Then \( \mu^* > 1 \) is determined by condition (70), and implies in an increase in the inflation rate \( \Pi^* = \mu^*/n > 1/n \). Since the equilibrium budget constraint is given by the equation \( c^o = (\Pi^*)^{-1}y - (\Pi^*)^{-1}c^y \), a higher inflation has the effect of making this line flatter (rotating it around the endowment point). Note that the analysis here resembles Figure 3.2 (An Expansionary Fiscal Policy Financed with an Income Tax). This is because the inflation tax is a distortionary tax (it makes the ‘cash’ good \( c^o \) more expensive relative to the ‘non-cash’ good \( c^y \)).

As with a change in relative prices, there will be substitution and wealth effects at work that determine the new position of the indifference curve. In Figure 9.1, I depict the new equilibrium by point B, which is associated with a lower demand for real money balances.

**Exercise 9.1:** Imagine instead that the government chooses to finance \( g \) with a lump-sum tax on the young (so that \( \mu = 1 \)). In this case, the equilibrium inflation rate is given by \( 1/n \). Show that the equilibrium budget constraint \( c^o + c^y/n = y - g \) must pass through point B in Figure 9.1. Explain why the representative young person prefers the lump-sum tax to the inflation tax. Bonus: explain why this is true of the initial old as well.

**Exercise 9.2:** In the model developed here, \( y \) is interpreted as an endowment of output, so that the real GDP (\( Nt\bar{y} \)) is unaffected by inflation. Imagine instead that \( y \) represents an endowment of time, so that \( c^y \) represents the consumption of leisure and \( (y - c^y) \) represents employment. Assume that \( (y - c^y) \) units of employment produces \( (y - c^y) \) units of current output. In this case, the real GDP is given by \( Nt(y - c^y) \). Explain how an increase in inflation can result in a decline in real GDP.

The analysis here suggests that the government’s ability to print new money does not allow it to create real wealth; it serves primarily as a way in which to extract wealth from those who value its money. Essentially, new money (created by the government) competes with old money (held by private agents) in a bid to purchase the economy’s output. A natural question to ask is whether there are any natural limits to this power. Does the government’s control of the money supply give it the power to extract resources without limit? Generally speaking, the answer is no. Let’s see why.
For the moment, let us take \( \mu \) as exogenous. Observe that the right-hand-side of equation (71) consists of two terms. The term \( [1 - \mu^{-1}] \) represents the inflation tax rate. The term \( q^D(y, \mu/n) \) represents the inflation tax base. Total tax revenue (seigniorage) is given by the tax rate multiplied by the tax base; i.e., \( S(\mu) \equiv (1 - 1/\mu)q^D(y, \mu/n) \).

How does seigniorage revenue depend on the money growth rate? We know one thing for sure: if \( \mu = 1 \) (a constant money supply), then \( S(1) = 0 \). In short, a zero tax rate implies zero taxes. We can be fairly confident of another thing: as \( \mu \to \infty \), then \( q^D \to 0 \). That is, as the money growth rate approaches infinity, the demand for real money balances approaches zero. The intuition for this is simple. Note that as \( \mu \to \infty \), the tax rate \( (1 - \mu^{-1}) \to 1 \) (i.e., the tax rate on money holdings is 100%). To put things another way, the real rate of return on money falls to zero as inflation heads off to infinity. As the return on money falls (i.e., as inflation rises), people are likely to demand less money (they substitute out of ‘cash’ goods into ‘non-cash’ goods). This has the effect of reducing the tax base. For very high rates of inflation, it is conceivable that \( q^D = 0 \), which would imply \( S(\infty) = 0 \).

On the one hand then, an increase in the money growth rate increases the
inflation tax rate and hence increases seigniorage revenue. On the other hand, the increase in inflation reduces the demand for money, thereby decreasing the tax base, thereby reducing seigniorage revenue. Whether seigniorage revenue rises or falls with \( \mu \) depends on the relative strength of these two effects. Figure 9.2 plots \( S(\mu) \) and demonstrates a case in which there is a clear limit to seigniorage as a revenue device (depicted by \( g_{\text{max}} \)). The shape of this function suggests that at very high rates of inflation, the government may actually be able to increase its seigniorage revenue by decreasing the inflation rate. This phenomenon is known as a Laffer effect.\(^{77}\)

Let us now return to assuming an exogenous level of \( g < g_{\text{max}} \) by supposing that \( \mu \) is determined endogenously by condition (71). If \( S(\mu) \) takes on the shape depicted in Figure 9.2, then our model implies that two equilibrium inflation rates are possible. One of these equilibria is associated with a low-inflation regime \( (\mu^*_L) \) and one is associated with a high-inflation regime \( (\mu^*_H) \). The analysis in Figure 9.1 suggests that the low-inflation regime has superior welfare properties.

**FIGURE 9.2**
Seigniorage and the Inflation Tax

Interpreting the Data

Over long periods of time, the power of seigniorage (the inflation tax) appears

\(^{77}\)Named after the American economist Art Laffer; see: http://en.wikipedia.org/wiki/Art_Laffer
to be used in most countries to extract a relatively small—but often significant—fraction of an economy’s resources for government purposes; see Table 9.1.

Our theory suggests a direct link between the relative size of government (say, as measured by $g$) and the fraction of GDP secured via seigniorage. This is because I have assumed that all government spending is financed via seigniorage. In reality, the correlation between government spending and seigniorage is not so tight. This can be explained by the fact that, for a given level of government spending, countries may choose alternative financing strategies (with some countries relying more heavily on seigniorage, and others relying more heavily on direct taxation).

Precisely what determines the finance decision is an interesting question. One hypothesis is that countries with relatively less-developed tax collection agencies may find the inflation tax an economical alternative. Another hypothesis is that countries with weak central banks are prone to political pressure from the fiscal authority to create money (thereby relieving the government’s need to draw on politically unpopular direct taxation).

Table 9.1
Seigniorage Revenue for Selected Countries (1960-1999)\textsuperscript{78}

It is interesting to note that even countries that do not on average over time depend heavily on seigniorage (like the United States) may, at times, find it expedient to collect seigniorage revenue during periods of fiscal crisis (e.g., an unusually heavy demand for government purchases). Periods of war constitute classic episodes that place unusually high (and presumably temporary) strain on the fiscal authority. During such episodes, it is invariably the case that the

fiscal authority tilts more heavily to money creation as a revenue device. The inevitable result is inflation.

Figure 9.3 plots the behavior of the inflation rate in the United States from 1890–2000. In the early part of the sample, the U.S. was on a gold standard. What this means is that government paper money was backed by the promise of redemption in gold (so that it was not technically fiat). 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, with the U.S. government promising to redeem other central banks’ holdings of dollars for gold at a fixed rate of $35 per ounce. Again, owing largely to the fiscal pressures brought on by the U.S.-Vietnam war (1963–75), the U.S. was compelled to abandon the gold standard (leading to a breakdown of the Bretton Woods system in 1971).

Figure 9.3 reveals an unmistakable link between periods of war and inflation (this pattern is one of the most robust in all of economic history). The theory developed above is consistent with this evidence: periods of unusually high government spending requirements are typically financed—at least in part—by turning to the printing press.

Source: www.j-bradford-delong.net/ Econ_Articles/ woodstock/ woodstock4.html
3. Money Superneutrality

In the previous chapter, we discussed the concept of money neutrality. Money is said to be **neutral** if an unanticipated one-time increase in the level of the money supply has no effect on real economic variables. In this section, I discuss a related—but conceptually distinct concept—call money superneutrality. Money is said to be **superneutral** if an increase in the money growth rate has no effect on real economic variables.

In this model economy, money is neutral in the long-run and may or may not be neutral in the short-run (depending on how the new money is injected into the economy). It turns out that money need not be superneutral even in the long-run. To see this, imagine that the government increases the money supply at some constant rate $\mu > 1$. Assume a constant population ($n = 1$) and that new money is injected by way of a lump-sum transfer $A_t$ to the old. In this case, the government’s budget constraint is given by:

$$ A_t = p_t(M_{t+1} - M_t)/N = [1 - 1/\mu] p_t M_t / N. $$

Define $a_t \equiv A_t(p_t N)$; i.e., the purchasing power of the money transfer accruing to each old agent. In a stationary (long-run) equilibrium, $a_t = a_{t+1} = a.$
From the perspective of a young agent, the budget constraint is now given by:

\[ c^y + \Pi c^o = y + \Pi a. \]

The solution to a young agent’s decision problem is given by a demand for real money balances \( q^D(y, \Pi, a) \).

**Exercise 9.3:** With the aid of a diagram, explain why \( q^D \) is an decreasing function of \( a \).

The market clearing condition implies that \( q^D(y, \Pi, a) = \frac{M_t}{N_p t} \) for all \( t \), so that \( \Pi^* = \mu \). Combining this with the government’s budget constraint yields:

\[ a^* = \left[ 1 - \frac{1}{\mu} \right] q^D(y, \mu, a^*). \]

This condition describes the equilibrium value of the transfer, \( a^* \). Note the similarity between this condition and (71). The difference can be explained by the fact that in (71), new money is used to finance government purchases \( g > 0 \); whereas here it is used to finance a ‘social security’ transfer, \( a \). In Figure 9.4, the initial equilibrium (with \( \mu = 1 \)) is depicted by point A, while the new (long-run) equilibrium (with \( \mu > 1 \)) is depicted by point B.

**FIGURE 9.4**
Money Non-Superneutrality
Exercise 9.4: Compare Figure 9.4 with Figure 3.3 (*A Lump-sum Transfer Financed with an Income Tax*) and discuss the similarities.

Exercise 9.5: In Chapter 4, we learned that the relative price of time-dated consumption can be interpreted as the real rate of interest. According to Figure 9.4, what effect does a policy of increasing the money supply via lump-sum transfers to the old have on the real rate of interest?

The example developed above demonstrates that money need not be superneutral, even in the long-run. But once again, our model suggests that this conclusion depends critically on the manner in which money is injected into the economy. In the example above, money is injected by way of lump-sum transfers (to the old). Imagine instead that new money is by way of a proportional transfer.

By a proportional transfer, I mean the following. Let $m_t$ denote the money held by a person when young. Then the amount of money this person holds increases in proportion to $m_t$ when old; i.e., $R^n m_t$. Note that this policy effectively pays interest on money; $R^n$ can be interpreted as the (gross) nominal interest rate.

To study the impact of this policy on behavior, write down the period budget constraints facing a representative young person:

\[ p_t c^y + m_t = y; \]
\[ p_{t+1} c^o = R^n m_t. \]

These may alternatively be expressed as:

\[ c^y + q = y; \]
\[ c^o = \left(\frac{R^n}{\Pi}\right)q; \]

where, recall, $q \equiv m_t/p_t$ and $\Pi \equiv p_{t+1}/p_t$.

Exercise 9.6: For a given $(R^n/\Pi)$, draw the budget line on a diagram and depict the demand for real money balances $q^D(y, R^n/\Pi)$. Explain how $R^n/\Pi$ can be interpreted as the real interest rate.

The next step involves formulated the government’s budget constraint. At the beginning of each period, the government must somehow come up with $(R^n - 1)M_t$ dollars (the net interest owing on its outstanding money supply). One way to do this would be by way of a tax. Another way would be to simply print up the new dollars needed, in which case:

\[ (R^n - 1)M_t = M_{t+1} - M_t; \]
\[ \Rightarrow R^n M_t = M_{t+1}. \]
Notice that this latter policy implies that the money supply will expand at the (gross) rate $R^n$.

Finally, let us look at the market-clearing conditions:

$$M_t = p_t N_t q^D(y, R^n / \Pi)$$ for all $t \geq 1$.

In a stationary (long-run) equilibrium, these conditions imply:

$$\frac{p_{t+1}}{p_t} = \Pi^* = \frac{M_{t+1}}{M_t} \frac{N_t}{N_{t+1}} = R^n / n.$$

**Exercise 9.7:** Demonstrate that $\Pi^* = R^n / n$ implies that the equilibrium budget line corresponds to the resource constraint, with $q^* = q^D(y, n)$.

The previous exercise demonstrates that this policy of expanding the money supply (to finance the interest on money holdings) is superneutral. While the inflation that results from this policy reduces the real return on money, offsetting this effect is the nominal interest that money now earns. In equilibrium, these two effects cancel out exactly—leaving the real return on money equal to $n$.

4. Monetary versus Fiscal Policy

*Inflation is always and everywhere a monetary phenomenon.*

– Nobel Laureate Milton Friedman

*Everything reminds Milton Friedman of the money supply.*

*Well, everything reminds me of sex, but I try to keep it out of my papers.*

– Nobel Laureate Robert Solow

To this point, I have not made any meaningful distinction between monetary and fiscal policy. In particular, we have thus far considered economies where a government that undertakes expenditures (in the form of either purchases or transfers) and then settles on some manner of financing these expenditures (e.g., by printing new money). Essentially, we assumed that the fiscal branch of the government was also in control of the money supply. In the context of many modern economies, however, there is a distinction to be made between monetary and fiscal policy. Let me explain.
The fiscal branch of the government (the fiscal authority) controls purchases and taxes; in nominal terms, we can denote these $G_t$ and $T_t$, respectively. The fiscal authority also controls the total level of government debt. For simplicity, assume that this debt takes the form of nominal bonds that mature in one period. Let $D_t$ denote the bonds that are due to mature (into cash) at the beginning of date $t$ and let $R^n_t$ denote the (gross) nominal interest rate paid on this debt. Then the government’s budget constraint takes the form:

$$G_t + R^n_t D_t = T_t + P_t + D_{t+1}; \quad (72)$$

where $P_t$ denotes profit from government-owned enterprises.

Contrary to popular belief, the monetary branch of the government (the monetary authority) is legally prohibited from injecting money into the economy by way of cash transfers (this power lies in the realm of the fiscal authority). In most economies, the monetary authority takes the form of a central bank that is, by varying degrees, independent of the fiscal authority.

If a central bank is unable to inject new money into the economy by way of transfers, then exactly how does new money find its way into the economy? The way this is done is through swaps of government money for government bonds. In other words, monetary policy determines how the total government debt in the hands of the public is divided between its interest-bearing (bonds) and non-interest-bearing (money) components. If we let $B_t$ denote the interest-bearing government bonds held by the public, then $M_t$ denotes the interest-bearing bonds held by the central bank (non-interest-bearing cash held by the public) and:

$$D_t = M_t + B_t. \quad (73)$$

Thus, at any point in time, the central bank holds assets (in the form of government bonds) and liabilities (in the form of cash, that was used to purchase the bonds—either directly from the government or from the public). Since bonds generally pay interest, while cash does not, the central bank makes a profit equal to:

$$P_t = (R^n_t - 1)M_t. \quad (74)$$

If the central bank is owned by the government, then this profit is returned to the government treasury. Thus, combining equations (72), (73) and (74), we can write:

$$G_t + R^n_t [M_t + B_t] = T_t + (R^n_t - 1)M_t + M_{t+1} + B_{t+1};$$

$$G_t + R^n_t B_t = T_t + B_{t+1} + [M_{t+1} - M_t]. \quad (75)$$

Equation (75) is just a more general version of the government budget constraint that I have employed throughout the text. In Chapter 3, we considered budget balance on a period-by-period basis, so that $G_t = T_t$ and $B_t = M_t = 0$. 

212
for all \( t \). In Chapter 5, we continued to assume \( M_t = 0 \), but allowed for the possibility of government debt. In the context of the two-period model studied there:

\[
\begin{align*}
g_1 + R_1 b_1 &= \tau_1 + b_2; \\
g_2 + R_2 b_2 &= \tau_2 + b_3;
\end{align*}
\]

where \( b_1 = 0 = b_3 \). In a monetary economy (an economy where government fiat money is valued), we saw in this chapter that the government has a third way to acquire purchasing power; i.e., by printing new money (assuming that the fiscal authority has power over the monetary authority). In Section 2 of this chapter, we considered the case in which \( B_t = T_t = 0 \), so that \( G_t = M_{t+1} - M_t \). In Section 3 of this chapter, we began by setting \( G_t = B_t = 0 \) and \( T_t = -A_t \), with \( A_t = M_{t+1} - M_t \). Later on in Section 3, we considered the case of interest-bearing government money. This latter case turns out to be equivalent to assuming \( G_t = T_t = M_t = 0 \) and \( B_{t+1} = R^n B_t \), with bonds serving as the economy’s money instrument.

Equation (75) calls into question Friedman’s assertion that inflation is always and everywhere a monetary phenomenon. Of course, there is a sense in which this statement must be true (for example, one might similarly declare that drowning is always and everywhere a watery phenomenon). But the deeper cause of inflation must lie in the conduct of fiscal policy and the power relationship that exists between the fiscal and monetary authorities of an economy. Throughout much of history and even today, the fiscal authority is usually the dominant agency. Fiscal authorities want to spend and need to tax (or issue debt) to finance their spending. The inflation tax is but one of many tax instruments that the fiscal authority would like to use. This is one of the main reasons for why many economists advocate central bank independence. The presumption here is that a ‘strong’ (independent) central bank is free to pursue policies that keep inflation under control.

5. 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.
1. Introduction

In Chapters 8 and 9, we looked at models that featured fiat money as the economy’s only asset. In reality, fiat money must compete with other assets that may conceivably be used in payments. These other assets may take the form of government bonds, which resemble fiat money in many ways, or in the form of private (intermediated) capital, such as your checking account.

The existence of alternative assets other than government fiat presents somewhat of a problem for monetary theorists. While fiat money is obviously useful as a payment instrument, it tends to possess a terrible rate of return (since inflation tends to be positive and since fiat money does not typically pay interest). One might expect (and indeed, history serves as a guide) that the private sector would be more than willing to supply privately-issued payment instruments to meet the demand for liquidity. Competition among liquidity-providers would likely ensure that the resulting payment instruments have very desirable properties (at least, absent legal restrictions that would prevent the private sector from designing optimal instruments). Theory suggests that government fiat money should continue to exist only to the extent that it offers a competitive rate of return, or if the private sector was somehow unable to replicate some of the desirable properties of fiat money. In reality, we observe fiat money coexisting with higher return assets (like interest-bearing checking accounts) that are also used in payments. This coexistence phenomenon remains somewhat of a puzzle.

One benefit of introducing multiple assets is that we can model more accurately the manner in which monetary policy is actually conducted. As I remarked in the previous chapter, monetary policy is usually implemented by way of debt swaps (government money for government bonds). If bonds dominate money in rate of return, then a monetary policy that alters the bond-money ratio can have real and long-lasting effects in the OLG model; these are discussed below.

2. An OLG Model with Capital

I extend the OLG model studied in the previous two chapters to allow for capital investment. As before, there is an initial old generation that has an exogenous endowment (which we can harmlessly set to zero). The young have an endowment \((y, 0)\), as before and value consumption when young and old \(u(c^y, c^o)\). The population of young agents grows exogenously according to \(N_t = \)
Imagine now that the young have access to a storage technology such that \( x \) units of investment today yields \( zf(x) \) units of output tomorrow (see Chapter 6). Here, you can interpret \( z > 0 \) an exogenous productivity parameter that affects the (expected) return to investment. Assume that more investment yields more output, so that \( f'(x) > 0 \); and that there are diminishing returns to capital investment, so that \( f''(x) < 0 \). As well, assume that \( f(0) = 0 \) (no investment means no return). Finally, assume that capital depreciates fully after it is utilized in production.

As in Chapter 6, the availability of a storage technology allows agents to smooth their consumption over time. This will be the case even in autarky. The autarkic allocation can be derived as follows. First, the initial old simply consume their endowment (zero). A representative young generation solves the following problem:

Choose \((\hat{c}^y, \hat{c}^o)\) to maximize \( u(c^y, c^o) \) subject to: \( c^o = zf(y - c^y) \);

where \( x = y - c^y \). The solution \((\hat{c}^y, \hat{c}^o)\) is described by two mathematical conditions:

\[
MRS(\hat{c}^y, \hat{c}^o) = zf'(y - \hat{c}^y);
\]
\[
\hat{c}^o = zf(y - \hat{c}^y);
\]

i.e., refer to Figure 6.3, where here we have \( MPK(\hat{x}, z) \equiv zf'(\hat{x}) \).

Exercise 10.1: Use the analysis discussed in Section 5 of Chapter 6 to argue that the equilibrium (gross) real rate of interest in a bond market (if such a market became available) would be given by \( \hat{R} = MPK(\hat{x}, z) \). Hint: note that this is a closed economy with a representative (young) agent, and that the lack of double coincidence precludes any intergenerational trade.

The analysis so far is just a rehash of the closed-economy analysis in Chapter 6. The next objective is to introduce fiat money into the economy and ask whether money can coexist with capital. The answer to this question is not so obvious. In the model described above, young agents can give up some of a good they value \( (c^y) \) to obtain another good of value \( (c^o) \). The way they accomplish this is by investing in capital. But if one can obtain \( c^o \) by investing in capital, why should there be a need to acquire fiat money (which is also used to acquire \( c^o \))? In this model, fiat money and capital are two assets that serve the same purpose (the acquisition of \( c^o \)). In Chapter 6, we studied a model with two

\( nN_{t-1} \), with \( N_0 > 0 \) given.
assets (a real bond and capital). We argued there that if these two assets are risk-free, then the principle of no-arbitrage implies that their returns should be the same (if both assets are to be willingly held by agents). As it turns out, the same principle will be at work here. In particular, if money and capital are both risk-free assets, then both assets must yield the same rate of return (if both are to be willingly held by agents). Conversely, if one asset dominates the other in rate of return, then the inferior asset will not be valued. One is then led to wonder whether the existence of private monetary instruments (capital) might not drive government fiat.

On Capital as Private Money

You may be wondering at this point about my reference to capital as a form of private money. How can a machine, a house, or inventory, be thought of as money? The answer is that while physical objects like these do not typically serve as a means of payment (with the possible exception of commodity monies, like gold or cigarettes), private debt instruments (in paper or electronic form) that are backed (collateralized) by such objects can and do serve as payment instruments. In the A-B-C model of private money studied early in Chapter 8, for example, money took the form of a private debt instrument backed by a promise of future output \( y_3 \).

You may be surprised to learn that, in fact, most of the money supply in any well-developed economy is in the form of private money. For the most part, this private money is created by the banking sector. In earlier times, banks would issue liabilities in the form of small denomination paper notes. These notes were redeemable in gold and backed by the bank’s capital.\(^1\) Nowadays, this money is primarily in electronic form (i.e., it exists as electronic book-keeping entries in your checking or savings account) and is redeemable in government cash. You undertake such redemptions every time you make a withdrawal from the bank or an ATM.

Economists distinguish between different forms of money by attaching particular labels to them. Government money, in the form of paper notes, is referred to as base money or high-powered money. Private money, in the form of electronic credits (redeemable for government money), is referred to as demand deposit money. The total money supply is defined to be the sum of base money and demand deposit money; this is labelled \( M_1 \).\(^2\)

In what follows then, whenever I speak of capital, I want you to think of private bank money. That is, think of the bank (or banking sector) as an agency that operates the investment technology on behalf of its owners (a young agent).

---

\(^1\)In the event of bankruptcy, for example, bank notes constituted a senior claim on the bank’s physical assets.

\(^2\)There are also broader definitions of money that include various forms of private debt that are not demandable but still commonly used in payments; e.g., \( M_2 \), \( M_3 \), etc.
The banking sector issues paper (or credits an account) that pays for (say) labor services (the endowment of the young) to construct new capital \( x \). In the future period, agents (who are now old) use their bank money to purchase output (produced by the maturing capital project).83

A Useful Simplification

In any model that features multiple assets, agents will be faced with two conceptually distinct problems: a saving decision and a portfolio allocation decision. We’ve already encountered this in Chapter 6, where we considered an economy with capital and bonds. We saw there that these two decisions can be decomposed into two steps. The first step is the portfolio allocation decision (undertaken to maximize wealth). The second step takes this maximum wealth as given and then determines the consumption-saving decision (undertaken to maximize utility).

We can simplify matters considerably by assuming that agents find it optimal to save their entire income. This would be the case, for example, if we assumed preferences of the form \( u(c^y, c^o) = c^o \) (i.e., young agents only care for consumption when old). This assumption allows us to focus on the portfolio allocation decision (since the saving decision is trivial). In this case, maximizing wealth is the same thing as maximizing utility. While not entirely realistic, this assumption allows us to isolate some basic economic forces that will continue to hold in a more general model in which this assumption is relaxed.

To see how this assumption makes life easier, let us reconsider the economy developed above, but with \( u(c^y, c^o) = c^o \). Draw yourself a diagram with a production possibilities frontier given by \( c^o = zf(y - c^y) \). Now draw some indifference curves (they will be horizontal and parallel lines in the commodity space). What is the solution to maximizing \( u(c^y, c^o) = c^o \) subject to \( c^o = zf(y - c^y) \)? Clearly, the solution is to set \( c^y = 0 \). This implies that \( c^o = zf(y) \) and \( x = y - c^y = y \). In other words, just save everything. End of story.

Exercise 10.2: Use a diagram to depict the solution to \( \max c^o \) subject to \( c^o = zf(y - c^y) \).

3. An OLG Model with Money and Capital

To begin our analysis, imagine that the government creates \( M \) dollars of fiat

\[\text{In reality, most private investment is financed by other debt instruments, like corporate debt and equity. But everything I say in this chapter will continue to hold if we assume that the component of private investment financed by private money is proportional to total investment.}\]
money and distributes this money to the initial old. Assume, for the moment, that \( M \) is held constant over time.

If money is to be valued (that is, if the young are to have an incentive to work for money), then the expected rate of return on money must be positive; i.e., \( \Pi_{t+1} < \infty \) (since the real rate of return on non-interest-bearing money is \( 1/\Pi_{t+1} \)). Furthermore, as money must compete with capital in the wealth portfolio of young agents, a simple no-arbitrage condition implies that the rate of return on these two assets must be equal (if they are both to be willingly held); i.e.,

\[
z f'(x_D^t) = 1/\Pi_{t+1}.
\]  

The condition above essentially describes the demand for investment \( x_D^t \) as a function of the expected return on money \( (1/\Pi_{t+1}) \) and the expected productivity of capital investment \( (z) \). The demand for real money balances is then given simply by \( q^D = y - x_D^t \) (where \( y \) denotes current income, all of which is saved).

You should be familiar with the theory of investment demand from Chapter 6. In particular, note that any force that serves to change the return on money (inflation rate) will lead to an asset-substitution (or portfolio rebalance) effect. For example, we see from (76) that an increase in \( \Pi_{t+1} \) will lead to an increase in \( x_D^t \). In other words, if agents perceive a reduction in the return on money, they will substitute some of their savings from money to capital investment. The substitution of money for capital is sometimes referred to as the **Tobin effect**.

Note that for this economy, the real GDP at date \( t \) is given by:

\[
Y_t = N_t y + N_{t-1} z f(x_{t-1}).
\]

Alternatively, we can use the expenditure approach to express GDP as:

\[
Y_t = N_{t-1} x_t^0 + N_t x_t.
\]

To this point, we have just assumed some \( \Pi_{t+1} < \infty \). The next step is to ask what determines inflation (and inflation expectations). Here, we focus on a stationary equilibrium in which \( \Pi_{t+1} = \Pi_{t+1} = \Pi \). In this case, investment demand (equation 76) is given by an \( x^D \) that satisfies:

\[
z f'(x_D^t) = \Pi^{-1}.
\]  

The demand for real money balances is then determined by \( q^D = y - x_D^t \). This will hold true for all \( t \geq 1 \).

---

84 Yes, I realize that this condition as stated is not exact. The proper condition is:

\[
E_t z_{t+1} f'(x_D^t) = E_t \Pi_{t+1}^{-1},
\]

where \( E_t \Pi_{t+1}^{-1} > 1/E_t \Pi_{t+1} \) by Jensen’s inequality.
The budget constraints facing (young) agents are given by:

\[ c_{t+1} = zf(x_t) + \Pi_{t+1}q_t; \]

or, in a stationary equilibrium:

\[ c^o = zf(x) + \Pi^{-1}q. \]

Once again, the initial old here are ‘special’ in that they do not have any initial endowment; i.e., \( x_0 = 0 \) so that \( zf(x_0) = 0 \). The initial old can then only consume to the extent that the young are willing to buy their money; i.e., \( c^o = q \).

The market clearing conditions, for all \( t \geq 1 \), are given by:

\[ M = p_tN_tq^D(\Pi^{-1}). \]

From this, it follows that the equilibrium inflation rate is given by:

\[ \Pi^* = 1/n. \] (78)

Since \( n > 0 \), it follows that \( \Pi^* < \infty \) (i.e., money is valued).

The equilibrium allocation can be derived as follows. First, by combining equations (77) and (78), the equilibrium level of capital investment is given by an \( x^* \) that solves:

\[ zf'(x^*) = n. \] (79)

From this, one can determine the equilibrium quantity of real money balances, \( q^* = y - x^* \). The initial old consume \( q^* \) and each young generation consume (when they are old) the amount \( c^{o*} = zf(x^*) + nq^* \). The monetary equilibrium is displayed in Figure 10.1.
Broad Money and the Money Multiplier

We can interpret the model above as one in which a private banking sector accepts deposits which are used to finance capital expenditure \( N_t x_t \). The nominal value of such deposits is given by \( p_t N_t x_t \), so that the total money supply (base money plus bank money) is given by:

\[
M_{1t} = M_t + p_t N_t x_t.
\]

From the market-clearing condition, we know that \( p_t = M_t/(N_t q_t) \). Substituting this into the expression above yields:

\[
M_{1t} = \left[ 1 + \frac{x_t}{q_t} \right] M_t.
\]  

The term in the square brackets above is called the money multiplier; i.e., the number by which we multiply the monetary base to find the total money supply. Observe that the money multiplier depends positively on the deposit-to-currency ratio \( x_t/q_t \).

In real economies, the broad money supply (M1) tends to be procyclical and appears to lead real GDP (i.e., an increase in \( M_{1t} \) tends to be followed
by an increase in $GDP_{t+1}$). Most of this movement is evidently attributable to changes in the money multiplier (as opposed to base money). Measures of broad money are also positively correlated with the price-level.

A simple (and common) interpretation of the data is that exogenous movements in the money supply somehow ‘stimulate’ real economic activity. The fact that changes in $M_1$ are correlated with changes in $GDP_{t+1}$ suggests that the direction of causality runs from money to output. I have warned you elsewhere not to interpret such correlations as evidence of causality. And indeed, there is good reason to believe here that the direction of causality in fact runs in reverse. Let me explain how this can be the case.

As we discussed in earlier chapters, an exogenous change in $z$ can be interpreted as an information (news) shock that occurs at date $t$ concerning the future return to capital spending. Good news leads agents to revise upward their forecast of $z$; bad news does the opposite. The actual realization may turn out to meet expectations—or it may not. I now order a specific expection, mistakes (after the fact) may be made, but agents are assumed to process information correctly. An alternative view is that expectations are not rational and that they simply move around for unexplained reasons (animal spirits).

In any case, let’s consider the following thought experiment. Imagine that the economy is as described above, with everyone expecting some given $z = z_L$. In this case, we can summarize the equilibrium by $q^*(z_L)$, with an initial price-level given by $p^*_1 = M/(N_tq^*(z_L))$. The initial equilibrium rate of inflation is given by $\frac{p^*_{t+1}}{p^*_t} = 1/n$.

Imagine now that at the beginning of date $t = 1$, agents suddenly (and unexpectedly) receive information that $z = z_H > z_L$ (good news). Assume that this is a permanent shock, so that the world continues to unfold as before, except for a higher $z$. In this case, equation (79) describes the new equilibrium level of investment $x^*(z_H) > x^*(z_L)$. That is, the expected return on money continues to be $n$ and the good news leads to a boom in capital expenditure.\(^8\) Given a fixed level of $y$, it follows that the demand for real money balances must fall: i.e., $q^*(z_H) < q^*(z_L)$. In short, people substitute out of (base) money and into capital (deposits). As the deposit-to-currency ratio rises, so does the money multiplier (and hence, the total money supply).

The effect of the shock is to cause a surprise increase in the current price-level: $p^*_1(z_H) > p^*_1(z_L)$. That is, while the supply of base money is fixed at $M$, a reduction in the demand for base money reduces its value; i.e., from the market-clearing condition, we have $p^*_1(z) = M/(N_tq^*(z))$. Note that the shock leads to a jump in the current price-level—and all future price-levels as well—so that future inflation remains unchanged (the level of prices jump up, but the rate of change in the future price-levels remains unaffected). The shock has

\(^{8}\)As I have remarked before, note that investment booms despite no obvious physical change in the environment. For this reason, an outside observer may be tempted to interpret such a shock as an ‘aggregate demand’ shock (an unexplained increase in capital spending).
no effect on current-period GDP (which at date 1, is given by $N_1y$). However, the expansion in current period investment leads to an increase in the future level of GDP. Thus, the current period increase in M1 is followed a higher future level of GDP.

To summarize then, a ‘good news’ shock leads to an upward revision in the forecasted return to capital spending. If there is competition in the banking sector, this will show up as an increase in the real interest rate paid on deposits. Households divert their savings away from cash and into deposits. The increase in deposits is used to finance additional capital expenditure. The increase in the deposit-to-currency ratio leads to an increase in the money multiplier and an expansion of broad money. The price level rises not because of the increase in broad money, but because of a reduction in the demand for base money. The increase in bank money today is followed by an increase in future GDP (as the new investment spending becomes productive capital). Of course, the direction of causality here is entirely reversed. It is the future shock to productivity that causes a contemporaneous increase in broad money; not the other way around.

Exercise 10.3: You should be able to draw a diagram that shows why an increase in $z$ leads to an improvement in welfare for the representative young generation. Explain why this ‘good news’ shock actually harms the initial old.

Rate of Return Dominance

One unattractive feature of the model developed above is that (base) money and capital earn the same real rate of return. In reality, base money appears to be dominated in rate of return by many other assets (with similar risk characteristics). One of the key challenges in monetary theory is to explain why this should be so. Unfortunately, a deep understanding of this phenomenon is not yet within the profession’s grasp.

One way to generate rate of return dominance is to simply assume that the government imposes various legal restrictions that are designed to enhance the value of its paper. An obvious example here is a reserve requirement that governments often impose on private banks. In the context of our model, a reserve requirement can be modeled by the restriction:

$$q_t \geq \theta x_t; \quad (81)$$

with $0 < \theta < 1$ representing the legal minimum currency-to-deposit ratio. Alternatively, one might just assume that the legal restriction takes the following simple form:

$$q_t \geq \sigma > 0. \quad (82)$$

In the monetary equilibrium studied above, private agents choose an optimal portfolio of money and capital ($q^*, x^*$). Suppose now that the government imposes a legal restriction in the form of (82). If $q^* > \sigma$, then the legal restriction
is non-binding (i.e., people willingly hold more than the minimum amount of cash required by law). On the other hand, the legal restriction will be binding if \( q^* < \sigma \). In this latter case, agents will have to reduce their capital holdings and substitute into money. Presumably, agents will want to hold as little cash as possible; i.e., \( q^0 = \sigma \) (so that \( x^0 = y - \sigma \)). The effect of such a legal restriction is displayed in Figure 10.2.

**FIGURE 10.2**
Legal Restriction and Rate of Return Dominance

\[
\begin{align*}
Slope &= -f'(y-\sigma) \\
Slope &= -n
\end{align*}
\]

From Figure 10.2, we see that the binding legal restriction has no effect on the equilibrium inflation rate. However, the restriction serves to artificially increase the demand for fiat money, so that resources are diverted away from capital investment and toward consumption (since the old use money to consume). The equilibrium rate of return on capital is now given by \( f'(y - \sigma) > n \), so that capital dominates money in rate of return. The binding legal restriction prevents the forces of arbitrage from equating these rates of return.

**Exercise 10.4:** Explain how a binding legal restriction for fiat money balances reduces the welfare of all young generations, but improves the welfare of the initial old generation.
4. Money, Bonds, and Capital

Let us now consider introducing another asset into our model economy in the form of a government bond. In some sense, fiat money is a type of government bond. In particular, it is a perpetual debt instrument (i.e., it has no fixed maturity date) that yields zero nominal interest.

In reality, government bonds come in a variety of forms and are distinguished from money primarily by the fact that bonds (usually) pay interest. To highlight this distinction, imagine that the government issues perpetual bonds that promise to pay interest (in the form of cash) every period. Assume that although bonds (like money) have no maturity date, they can (like money) be used in making payments.

Government policy works as follows. Imagine that the government creates a given level of nominal debt $D$, which it holds fixed forever. The government endows this debt, in the form of money and bonds, to the initial old: $D = M + B$. The interest on bonds $(R^n - 1)B$ is financed by a lump-sum tax $T$ on the old. There is a (binding) legal restriction in the form of (82).

The choice problem of an initial old person is easy: just sell $M + B$ for as many goods as possible. The choice problem for a representative young person is a little more complicated. First, suppose that the government sets $R^n > 1$. What this means is that bonds dominate money in rate of return. It follows that the legal restriction (82) will bind, so $q^* = \sigma > 0$. Next, a simple no-arbitrage argument suggests that the demand for investment is determined by:

$$zf^n(x^D) = \frac{R^n}{\Pi} > \frac{1}{\Pi}.$$ 

With the demand for investment and money determined in this way, the (real) demand for government bonds is given by $b^D = y - \sigma - x^D$.

Consumption (when old) is given by:

$$c^o = zf(x^D) + \frac{R^n}{\Pi}b^D + \frac{1}{\Pi}\sigma - \tau,$$

where $\tau$ represents the (real) tax paid by old agents.

To simplify slightly, let’s assume a constant population. The market-clearing condition in the money-goods market can be used to derive the equilibrium price-level (and inflation rate):

$$p^* = \frac{M}{N\sigma} \Rightarrow \Pi^* = 1.$$ 

Now, let’s look at the government’s budget constraint $T = (R^n - 1)B$, which, in real per capita terms, can be written as:

$$\tau = \Pi^{-1}(R^n - 1)b.$$
where $\tau \equiv T/(Np)$ and $b \equiv B(N/p)$. Combining this with the old-age budget constraint above yields:

$$
e^o = zf(x) + b + \sigma. \tag{83}$$

But since $b = B(N/p)$ and $p^* = M/(N\sigma)$, it follows that $b^* = (B/M)\sigma$. Therefore, $x^* = y - \sigma - b^* = y - [1 + B/M] \sigma$, which implies an equilibrium real (and nominal) interest rate:

$$R^* = zf^*(y - [1 + B/M] \sigma).$$

The equation above is interesting because it reveals how (in this model) the division of a given amount of government debt $D$ between its interest-bearing ($B$) and non-interest-bearing ($M$) components can affect real economic variables. You may recall that such an operation is what characterizes monetary policy from fiscal policy. One can think of an increase in $B/M$ as an open-market operation involving a sale of bonds (on the part of the central bank to the public) in exchange for government cash (a contractionary monetary policy). Likewise, one can think of a decrease in $B/M$ as an open-market operation involving a purchase of bonds for government cash (an expansionary monetary policy).

According to equation (83), a contractionary monetary policy has the effect of increasing the real (and nominal) interest rate. The higher interest rate is associated with a contraction in real capital spending. The intuition is as follows. Note that this contractionary monetary policy has the effect of increasing the supply of government bonds to the public. These bonds must compete with capital in the wealth portfolios of agents. The increased supply of bonds reduces the price of bonds, thereby increasing their yield. Higher bond yields attract saving away from capital and into bonds. Notice too that the implied reduction in the base money supply leads to a reduction in the equilibrium price-level.

A Liquidity Trap

Perhaps you’ve heard of the so-called liquidity trap. As far as I can tell, almost no one knows what it means, although everyone seems to agree that it sounds scary and should be avoided at all costs.

A liquidity trap occurs when the rate of return differential on (government) money and (government) bonds is zero. In this case, money and bonds are likely to be viewed as (close to) perfect substitutes by private agents (e.g., chartered banks). Since monetary policy is conducted by way of swaps of money for (government) bonds, such swaps can have no effect if the two assets are essentially identical. Monetary policy can no longer have any real effect on the economy.

Since the nominal interest rate paid on cash is zero, a liquidity trap arises when the nominal interest rate (on government bonds) falls to zero (or close
to zero). The monetary authority cannot make the nominal interest rate fall below zero (i.e., \( R^n < 1 \) is not possible). An expansionary monetary policy in this case (an decrease in \( B/M \)) would simply lead private agents to substitute (zero interest) bonds with (zero interest) cash in their wealth portfolios. In the financial pages of newspapers, this effect would be described as agents (in particular, banks) as ‘hoarding’ cash instead of spending or lending it.

What forces are responsible for generating a liquidity trap? Another way to phrase this question is to ask what forces are responsible for generating a low-interest rate environment. The model developed above suggests that one such force may be a pessimistic outlook regarding the expected return to domestic capital spending (i.e., a low value for \( z \)). For a given rate of inflation, pessimism of this sort (whether warranted or not) will put downward pressure on the real (and nominal) interest rate. If the interest rate falls low enough, the monetary authority will find itself in a liquidity trap.

Note that a liquidity trap is neither a good nor bad thing in itself. All it means is that the monetary authority has no way of affecting real economic variables by way of conventional swaps of money for bonds. This may be a good thing, if the monetary authority is crazy, for example. It also does not mean that the government cannot affect the economy in other ways. It is easy to demonstrate, for example, that the government can still use fiscal policy to influence the inflation rate, even if the nominal interest rate is equal to zero.

**Exercise 10.5:** Using the model developed above, assume that the equilibrium nominal interest rate is zero (i.e., that \( R^{n*} = 1 \)). Demonstrate that if this is the case, then the legal restriction \( q \geq \sigma \) cannot bind (and explain why). Characterize the general equilibrium level of investment and show how the equilibrium in no way depends on \( B/M \).

5. Summary

Incomplete.
CHAPTER 11

International Monetary Systems

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 $1/0.80 = $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 were apparently surprised by the subsequent ‘excessive’ volatility in exchange rate behavior.

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).

86 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 consists of USD.
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.

2. The Nominal Exchange Rate in a Free Market

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 $p^a_t$ and $p^b_t$ 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:

$$ p^a_t = e_t p^b_t; \quad (84) $$

where $e_t$ denotes the price of country $b$ money measured in units of country $a$ money (i.e., the nominal exchange rate). Equation (84) 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., \( q^D_a + q^D_b = y \). The rate of return on each money is inversely related to the inflation rate in each country; i.e.,

\[
\Pi^a = \frac{p_{t+1}^a}{p_t^a} \quad \text{and} \quad \Pi^b = \frac{p_{t+1}^b}{p_t^b}.
\]

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.,\(^87\)

\[
\Pi^a = \Pi^b = \Pi.
\]

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:

\[
\frac{M^a}{p_t^a} + \frac{M^b}{p_t^b} = 2N \left( q^D_a + q^D_b \right).
\]

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 \( q^D_a + q^D_b = y \), this equation may be rewritten as:

\[
M^a + e_t M^b = p_t^a 2N y.
\]

Equation (85) 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

\(^{87}\)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.
determine the exchange rate (and price-level), we need another equation. How-
ever, economic theory does not deliver any further restrictions in this environ-
ment. In other words, there are no economic fundamentals that determine the
nominal exchange rate. **Any** nominal exchange rate $e_t$ (together with a corre-
sponding price-level that satisfies 85) 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 equilib-
rium nominal exchange rate between two fiat currencies is driven **entirely by**
market speculation.

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 devel-
oping 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 intrinsi-
cally 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 (85), 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, $p^a$ now represents the price of output measured in Lauriers and $p^b$ 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, $p^a = 4p^b$. 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 (85) constitutes one equation in the one unknown; i.e., $p^a$. Thus, a government policy that fixes the nominal exchange rate implies that market forces can determine a unique price-level $p^a$ (with $p^b = 0.25p^a$).

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?

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.\footnote{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}
change 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.\(^8\)

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 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 expen-

---

ditures with money creation, then the government budget constraint implies:
\[ p_t N g = M_t^a - M_{t-1}^a; \]
\[ = \left[ 1 - \frac{1}{\mu_t^a} \right] M_t^a. \]

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:
\[ g = \left[ 1 - \frac{1}{\mu_a} \right] \frac{M_t^a}{Np_t^a}; \]
\[ = \left[ 1 - \frac{1}{\mu_a} \right] y. \]

Thus, the domestic inflation rate that would be required to finance the expenditure level \( g \) is given by:
\[ \mu_a = \frac{1}{1 - \frac{2}{y}} > 1. \] (86)

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:
\[ p_t N g = M_t - M_{t-1}; \]
\[ = \left[ 1 - \frac{1}{\mu} \right] M_t; \]
where \( M_t = M_t^a + eM_t^b \) 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:
\[ M_t = p_t N (q^a + q^b); \]
\[ = p_t N 2y; \]
(since $q^a = q^b = y$ here). Now, substitute this money-market clearing condition into the government budget constraint:

$$g = \left[ 1 - \frac{1}{\mu} \right] \frac{M_t}{Np_t};$$

$$\quad = \left[ 1 - \frac{1}{\mu} \right] 2y.$$ 

In this case, the inflation rate that would be required to finance the expenditure level $g$ is given by:

$$\mu = \frac{1}{1 - \frac{1}{2} y}.$$  \hspace{1cm} (87)

Now, compare equations (86) and (87). 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 (87) instead of equation (86)).

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 (87) 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.

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 fol-
lowing scenario. Imagine that Canada and the U.S. have entered into a bilateral
fixed exchange rate agreement. Now, imagine that participants in the foreign ex-
change (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 ex-
change 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

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.\footnote{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.} 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.

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).

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.
3. The Nominal Exchange Rate Under 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.91 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’ (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

---

91In the late 1970s, the Bank of America wanted to offer deposits denominated in Japanese yen, but was officially discouraged from doing so.
consider different population growth rates $n^i$ and different money supply growth rates $\mu^i$ 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.,

$$M^a_t = p^a_t N^a_t y; \quad M^b_t = p^b_t N^b_t y.$$  

With domestic price-levels determined in this way, the equilibrium exchange is determined (by the LOP) as:

$$e^a_t = \frac{p^b_t}{p^a_t} = \frac{M^a_t}{M^b_t} \frac{N^b_t}{N^a_t}.$$  

(88)

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:

$$\frac{e_{t+1}}{e_t} = \frac{\Pi^a}{\Pi^b} = \frac{\mu^a n^b}{\mu^b n^a}.$$  

(89)

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 (88) 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 (89) 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 $e_{t+1}/e_t < 1$).
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 (89) 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.

3. 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’ alter-
native would be a multilateral fixed exchange rate arrangement, like Bretton-
Woods (sans foreign currency controls). Under this scenario, different cur-
rencies 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 im-
portantly, 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 coun-
tries 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 cur-
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 mone-
tary 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 pri-
vate 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 counter-
feiting (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.92 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.

92 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.
Appendix 11.A
Nominal Exchange Rate Indeterminacy and Sunspots

The following constitutes a simplified version of the model presented in Manuelli and Peck (1990). Time is discrete and the horizon is infinite; \( t = 1, 2, ..., \infty \). There are two economies, each of which are populated by a sequence of two-period-lived overlapping generations (with an initial old generation). Each economy features a constant population of \( 2N \) agents. A representative young agent has preferences defined only over old-age consumption, \( c_{t+1} \). Preferences are represented by:

\[ E_t u(c_{t+1}), \]

where \( u'' \leq 0 < u' \) with \( \lim_{c \to 0} u'(c) = \infty \) and \( E_t \) is an expectations operator conditional on information available at date \( t \). Young agents have a nonstorable endowment \( y_t > 0 \). This endowment follows an exogenous stochastic process.

Assume that the two countries both receive the same endowment shock. Let \( (2M, 2M^*) \) denote the fixed stocks of domestic and foreign money, respectively (with each money endowed evenly among the initial old in both economies). Let \( (v_t, v^*_t) \) denote the value of domestic and foreign currency, respectively, measured in units of output. Likewise, let \( (q_t, q^*_t) \) denote the quantity of real money balances of the domestic and foreign money held by a domestic agent. The choice problem facing a representative young agent at date \( t \) can be stated as follows:

\[
\max E_t u(c_{t+1})
\]

subject to:

\[
y_t = q_t + q^*_t;
\]

\[
R_{t+1} q_t + R^*_t q^*_t \geq c_{t+1};
\]

where \( R_{t+1} = (v_{t+1}/v_t) \) and \( R^*_t = (v^*_{t+1}/v^*_t) \). The initial old trivially consume \( c_1 = [v_1 M + v^*_1 M^*]/N \).

Note that since agents trivially save their entire endowment, their choice problem essentially boils down to a portfolio decision; i.e.,

\[
\max E_t u \left( R^*_{t+1} y_t + \left[ R_{t+1} - R^*_t \right] q_t \right),
\]

so that \( q^P_t \) satisfies:

\[
E_t \left[ R_{t+1} - R^*_t \right] u' \left( R^*_{t+1} y_t + \left[ R_{t+1} - R^*_t \right] q^P_t \right) = 0.
\]

Note further that since preferences and endowments are identical across countries, consumption will also be the same. If the exchange rate is constant (or if agents are risk-neutral), there is some indeterminacy in the optimal exchange rate.
portfolio but all such portfolios buy the same (expected) consumption. If the exchange rate fluctuates, then there is a unique optimal portfolio (if agents are risk-averse) so, again, consumption is the same across countries. Since all equilibria involve no net trade of goods between countries, we can without loss focus on symmetric equilibria where agents in each country hold the same portfolio (so that we need not introduce notation to distinguish between domestic and foreign agents). In this case, we have:

\[
\begin{align*}
  v_t M &= N q_t^D; \\
  v_t^* M^* &= N q_t^{*D}.
\end{align*}
\]

For simplicity, let us normalize such that \( M = M^* = N = 1 \), so that \( q_t^D = v_t \) and \( q_t^{*D} = v_t^* \). Now, from the budget constraint:

\[
\begin{align*}
  c_{t+1} &= R_{t+1} q_t^D + R_{t+1}^* q_t^{*D} \\
  &= R_{t+1} v_t + R_{t+1}^* v_t^* \\
  &= v_{t+1} + v_{t+1}^*.
\end{align*}
\]

The relevant money market clearing condition is given by:

\[
v_t M + v_t^* M^* = N y_t \quad \forall t\;
\]

Since this condition must hold at all dates, we have:

\[
\frac{v_{t+1} + v_{t+1}^*}{v_t + v_t^*} = \frac{y_{t+1}}{y_t}.
\]

The equation above represents the (gross) real rate of return on a diversified (symmetric) portfolio (i.e., in equilibrium, each young agent purchases one unit of domestic money and one unit of foreign money). Combining the money market clearing condition with the budget constraint reveals that \( c_{t+1} = y_{t+1} \); so we have the following restriction on equilibrium rates of return:

\[
E_t \left[ R_{t+1} - R_{t+1}^* \right] u'(y_{t+1}) = 0. \tag{90}
\]

This equation may alternatively be written as:

\[
E_t \left[ R_{t+1} - R_{t+1}^* \right] E_t u'(y_{t+1}) + cov_t \left( \left[ R_{t+1} - R_{t+1}^* \right], u'(y_{t+1}) \right) = 0. \tag{91}
\]

### Constant Exchange Rate Equilibria

Assume that \( y_t = y \), so that there is no ‘fundamental’ uncertainty. This restriction does not rule out the possibility of randomness in ‘non-fundamentals.’ In this section, however, assume that \( R_t \) and \( R_t^* \) are deterministic. In this case, condition (90) reduces to:

\[
R_{t+1} = R_{t+1}^*.
\]

247
Let $e_t$ denote the value of foreign money measured in units of domestic money, so that $v_t^* = e_t v_t$. Then this latter equation can be written as:

$$R^* + 1 = \left( \frac{e_{t+1}}{e_t} \right) R_{t+1};$$

which implies:

$$e_{t+1} = e_t = e.$$

In other words, in the absence of any uncertainty, the equilibrium exchange rate must be constant. From the money market clearing condition,

$$v_t + v_t^* = y;$$

$$(1 + e_t)v_t = y;$$

so that the (common) equilibrium rate of return on money is given by:

$$R_{t+1} = R^*_{t+1} = \frac{(1 + e_{t+1})y}{(1 + e_{t})y} = 1.$$

Note that since both currencies earn the same rate of return $R = R^* = 1$, the optimal portfolio decision at the individual level is indeterminate. That is, both currencies are literally perfect substitutes; since each is as good as the other, individuals are willing to hold any combination of $(q, q^*)$ such that $q + q^* = y$ (note: here, $y$ corresponds to the demand for total real balances). Since $v_t = q_t$, we can rewrite the money market clearing condition as:

$$(1 + e)q = y.$$

But since $q$ is indeterminate, so is $e$. In other words, any $e$ satisfying $0 < e < \infty$ is consistent with an equilibrium in which both currencies are willingly held. This is just the Kareken and Wallace (1981) indeterminacy result.

Imagine now that $y_t$ follows a stochastic process. The question I ask here is whether $e_t = e$ is still an equilibrium. To see whether this is the case, assume that $e_t = e$. In this case, $R_{t+1} = R^*_{t+1}$; which obviously implies $E_t [R_{t+1} - R^*_{t+1}] = 0$. In this scenario, the (common) equilibrium rate of return on money is given by:

$$R_{t+1} = R^*_{t+1} = \frac{y_{t+1}}{y_t}.$$

Of course, this rate of return is random, so that individuals attain an expected utility $E_t u(y_{t+1})$. From the money market clearing condition, we have:

$$(1 + e)v_t = y_t.$$

Once again, the two currencies are viewed as perfect substitutes. Thus, for a fixed $e$, individuals are perfectly willing to hold $q_t = v_t$ and $q^*_t = v^*_t$ units of
domestic and foreign money (respectively) in their portfolios, where
\[ v_t = \left( \frac{1}{1+e} \right) y_t; \]
\[ v_t^* = \left( \frac{e}{1+e} \right) y_t. \]

In this equilibrium, both the domestic and foreign price-levels fluctuate in exactly the same way. The rate of return on money fluctuates, but since the exchange rate is fixed, the returns on both monies fluctuate in exactly the same way. Thus, in this model, changing fundamentals need not induce fluctuations in the equilibrium exchange rate. As a corollary, it follows that bi-lateral fixed exchange rate system—or a currency union—is consistent with efficiency.

Volatile Exchange Rate Equilibria with Constant Fundamentals

Let us return now to the case in which there is no ‘fundamental’ uncertainty, so that \( y_t = y \). Now, since \( u'(y) \) is a constant, we can write the restriction (90) as:
\[ E_t \left[ R_t + R_t^* \right] = 0; \]
\[ E_t \left[ \frac{v_t+1}{v_t} - \frac{e_t+1}{e_t} \frac{v_t+1}{v_t} \right] = 0; \]
which reduces to:
\[ E_t e_t v_t+1 = e_t E_t v_t+1. \]

From the money market clearing condition, \( v_t + v_t^* = y \); or \( (1+e_t)v_t = y \). Recall, as well, that \( q_t = v_t \) and \( q_t^* = v_t^* \). Now, let’s follow Manelli and Peck (1990), and define the variable \( \alpha_t \) by:
\[ v_t = \alpha_t y. \]

Here, \( \alpha_t \) has the interpretation of being the fraction of one’s (real) saving allocated to the domestic currency. We want to restrict attention to situations where both currencies are always valued; i.e.,
\[ 0 < \alpha_t < 1. \]

Now, use this definition and the money market clearing condition to write \( e_t = (1 - \alpha_t)/\alpha_t \). Substituting this into our earlier restriction:
\[ E_t e_t v_t+1 = e_t E_t v_t+1; \]
\[ E_t \left( \frac{1 - \alpha_t+1}{\alpha_t+1} \right) \alpha_t+1 y = \left( \frac{1 - \alpha_t}{\alpha_t} \right) E_t \alpha_t+1 y; \]
\[ \alpha_t E_t (1 - \alpha_t+1) = (1 - \alpha_t) E_t \alpha_t+1; \]
which reduces to:

\[ \alpha_t = E_t \alpha_{t+1}. \]  

(92)

**Proposition:** (Manuelli and Peck, 1990). Any stochastic process \( \{\alpha_t\}_{t=1}^{\infty} \) satisfying \( 0 < \alpha_t < 1 \) and condition (92) constitutes an equilibrium (with both monies valued); with \( v_t = q_t = \alpha_t y \); \( v_t^* = q_t^* = (1 - \alpha_t) y \) and \( e_t = (1 - \alpha_t)/\alpha_t \).

**Remark:** Note that (92) implies that each country’s respective price-level follows a martingale (with the price-level in each country being perfectly negatively correlated with each other); in particular, it does not imply that the exchange rate itself is a martingale. In fact, the theory predicts that \((1 + e_t)^{-1}\) follows a martingale; i.e., one can rewrite (92) as:

\[ (1 + e_t)^{-1} = E_t (1 + e_{t+1})^{-1}. \]

**Example**

Let \( \{\alpha_t\}_{t=1}^{\infty} \) follow the process:

\[ \alpha_{t+1} = \alpha_t + \eta_{t+1}, \]

where \( \eta_{t+1} \sim U[-\kappa_t, \kappa_t] \) and \( \kappa_t \equiv \min\{\alpha_t, 1 - \alpha_t\} \). Note that \( E[\eta_{t+1} | \alpha_t] = 0 \) so that \( E[\alpha_{t+1} | \alpha_t] = \alpha_t \), as required (note: we are free to choose \( 0 < \alpha_t < 1 \) arbitrarily).
Appendix 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 $q_i^t$ 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:

$$M^a = p_i^a N(y + q_i^a);$$
$$M^b = p_i^b N(y + q_i^b).$$

The nominal exchange rate in this case is given by:

$$e_t = \frac{p^a_t}{p^b_t} = \frac{M^a}{M^b} \left( \frac{y + q^b_t}{y + q^a_t} \right) = \frac{M^a}{M^b} \left( \frac{2y - q^a_t}{y + q^a_t} \right).$$

This condition constitutes one equation in the two unknowns: $e_t$ and $q^a_t$. Hence, the exchange rate is indeterminate and may therefore fluctuate solely on the ‘whim’ of international currency traders (i.e., via their choice of $q^a_t$). 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 $b$. 

251
Appendix 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. As we saw during the recent Enron scandal, the market reacts quickly and ruthlessly when it gets a whiff of financial shenanigans.

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

---

95 For more on the Enron fiasco, see: www.washingtonpost.com/wp-dyn/business/specials/energy/enron/
nature of government policy).

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 (in my view) 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.
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 do not simply imitate their more advanced 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.

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. As is the case today, 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.

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, 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.

96These examples are taken from: Mokyr, Joel (1990). The Levers of Riches: Technological Creativity and Economic Progress, Oxford University Press, New York.
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.”97 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.

The Renaissance and Baroque Period (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.

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 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.
4. 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:

\[ Y_t = F(K, N_t), \]

(93)

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 12.1.
The technology for producing people is expressed as follows. First, assume that there is an exogenous birth rate $b > 0$. This assumption captures 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., $n_t = b - m_t$. Then it is clear that the population growth rate is an increasing function of living standards; a relation that we can write as:

$$n_t = n(y_t),$$

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:

$$N_{t+1} = [1 + n(y_t)]N_t.$$  \hspace{1cm} (95)

Note that when $y_t = y^*$ (in Figure 12.2), the net population growth rate is equal to zero (the birth rate is equal to the mortality rate) and the population stays constant.
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 12.2. Thus, initially at least, per capita incomes are above subsistence levels.

According to Figure 12.2, 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 12.1, 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 12.2 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 12.1, 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 even-
tually (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 12.3 depicts these transition dynamics.

FIGURE 12.3
Transition Dynamics

Exercise 12.1: Using a diagram similar to Figure 12.3, describe the dynamics that result when the initial population is such that $N_0 > N^*$. 

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 12.2: Using a diagram similar to Figure 12.3, 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.

An Improvement in Health Conditions

The number one cause of death in the history of mankind has not been war, but disease.\(^98\) 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 12.4. The effect of such a change is to immediately reduce mortality rates which, according to Malthusian 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 12.4. 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 12.4**
An Improvement in Health Technology

Exercise 12.3: 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 12.4, describe the dynamics for per capita income in the Malthusian model when the economy is subject to a transitory increase in the mortality rate.

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 Malthusian model has a difficult time accounting for the sustained in-
crease 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 12.2) 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 Malthu-
sian 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 de-
scription. 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 ac-
count 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 im-
provements prior to 1800 manifested themselves primarily in the form of larger
populations (and total output), with only modest improvements in per capita
incomes.

5. Fertility Choice

The Malthusian model does not actually model the fertility choices that
households make. The model simply assumes that the husband and wife just
automatically create children. Perhaps children are simply the by-product of un-
controllable 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 12.4:** 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:99

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 \( t = 0, 1, 2, \ldots, \infty \). 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:

\[
c_t(2) = n_{t-1} \theta w_t.
\]  

(96)

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:

\[
c_t(1) = (1 - \theta)w_t - n_t.
\]  

(97)

By substituting equation (97) into equation (96), with the latter equation updated one period, we can derive the following lifetime budget constraint for a representative young person:

\[
c_t(1) + \frac{c_{t+1}(2)}{\theta w_{t+1}} = (1 - \theta)w_t.
\]  

(98)

Equation (98) 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 12.5 displays the optimal choice for a given pattern of wages \((w_t, w_{t+1})\).
Figure 12.5 makes clear the analog between the savings decision analyzed in earlier chapters 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., $n_D^t = n^D(w_t, w_{t+1})$.

**Exercise 12.5:** 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 (93). 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 2.1, 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 12.6:** 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 define a function $n^*_t = \phi(N_t)$. A stable steady state will exist if $\phi'(N) < 0$, so let us make this assumption here. This condition asserts that the equilibrium population growth rate is a decreasing function of population size; see Figure 12.6.

![Equilibrium Population Dynamics](image)

**FIGURE 12.6**

Equilibrium Population Dynamics

---

100 The function $\phi$ is defined implicitly by:

$$\phi(N_t) = n^D(MPL(N_t), MPL(\phi(N_t))N_t).$$

101 I am pretty sure that for sufficiently large populations, the function $\phi$ must eventually decline with population since land is in fixed supply.
In Figure 12.6, 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. 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 12.5), the steady state consumption levels must satisfy:

$$\frac{1}{\beta} c^*(1) = \theta w^* = \theta MPL(N^*) .$$

Now, assume that the share parameter $\theta$ 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:

$$MPL(N^*) = \frac{1}{\theta \beta} = w^* .$$

Condition (99) tells us that the steady state population $N^*$ is determined by the nature of the production technology ($MPL$) and the nature of preferences ($\beta$).

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 $\phi$ in Figure 12.6 shifts upward). From equation (99), 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 12.5). The increasing population, however, puts downward pressure on the wage until it eventually falls to its initial value $w^*$. From equation (99), we see that the long-run wage rate depends only on $\beta$ 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

---

102 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.
land. But if land accounts for a relatively small fraction of total output, then
the impact on per capita GDP will be small.

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 1980.103 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. Summary

Incomplete.

103 http://nhs.needham.k12.ma.us/cur/kane98/kanep2/chinas1kid/dcva2.html
CHAPTER 13

Modern Economic Development

1. Introduction

In the previous chapter, we saw that despite the fact of technological progress throughout the ages, material living 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 and Pomeranz argue that living standards across countries in Europe, China, and the Indian subcontinent were roughly comparable in 1800. Parente and Prescott 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), Hargrave’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 9 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?  Clearly, something else other than just technological progress must

---

107 While populations did rise in the west, total income rose even faster.
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 13.1.

Table 13.1

<table>
<thead>
<tr>
<th>Year</th>
<th>West (1990 US$)</th>
<th>East (1990 US$)</th>
<th>West/East</th>
</tr>
</thead>
<tbody>
<tr>
<td>1820</td>
<td>1,140</td>
<td>540</td>
<td>2.1</td>
</tr>
<tr>
<td>1870</td>
<td>1,880</td>
<td>560</td>
<td>3.3</td>
</tr>
<tr>
<td>1900</td>
<td>2,870</td>
<td>580</td>
<td>4.2</td>
</tr>
<tr>
<td>1913</td>
<td>3,590</td>
<td>740</td>
<td>4.8</td>
</tr>
<tr>
<td>1950</td>
<td>5,450</td>
<td>727</td>
<td>7.5</td>
</tr>
<tr>
<td>1973</td>
<td>10,930</td>
<td>1,670</td>
<td>6.5</td>
</tr>
<tr>
<td>1989</td>
<td>13,980</td>
<td>2,970</td>
<td>4.7</td>
</tr>
<tr>
<td>1992</td>
<td>13,790</td>
<td>3,240</td>
<td>4.3</td>
</tr>
</tbody>
</table>

The data in Table 13.1 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 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 13.1 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 13.1. 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 13.1
Real per Capita GDP Relative to the United States
Selected Countries
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.\textsuperscript{108} 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).

I remarked in Chapter 12 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 12 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.\textsuperscript{109}

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.,

\[ Y_t = F(K_t, N_t), \]  

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

---


\textsuperscript{109} Of course, this does not explain why the institutional environments should differ the way that they do.
according to:

\[ K_{t+1} = K_t + X_t - \delta K_t. \tag{101} \]

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.,

\[ S_t = \sigma Y_t, \tag{102} \]

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:

\[ N_{t+1} = (1 + n)N_t, \tag{103} \]

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 (100), 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 (102) through by \( N_t \), we can rewrite it in per capita terms:

\[ s_t = \sigma y_t; \tag{104} \]

\[ = \sigma f(k_t). \]

Now, take equation (101) and rewrite it in the following way:

\[ \frac{N_{t+1} K_{t+1}}{N_{t+1} N_t} = \frac{K_t}{N_t} + \frac{X_t}{N_t} - \frac{\delta K_t}{N_t}. \]
Using equation (103), we can then express this relation as:

\[(1 + n)k_{t+1} = (1 - \delta)k_t + x_t,\]  

(105)

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 (104) and (105) to derive:

\[(1 + n)k_{t+1} = (1 - \delta)k_t + \sigma f(k_t).\]  

(106)

For any initial condition \(k_0\), equation (106) completely describes the dynamics of the Solow growth model. In particular, given some \(k_0\), we can use equation (106) to calculate \(k_1 = (1 + n)^{-1}[(1 - \delta)k_0 + f(k_0)]\). Then, knowing \(k_1\), we can calculate \(k_2 = (1 + n)^{-1}[(1 - \delta)k_1 + f(k_1)]\), 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 (106) is depicted graphically in Figure 13.2.

\[\text{FIGURE 13.2} \]
\text{Dynamics in the Solow Model}

As in the Malthus model, we see from Figure 13.2 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 \(k_t \uparrow k^*\) and
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 13.1). 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.

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 (106) that the steady state capital-labor ratio satisfies:

\[
\sigma f(k^*) = (n + \delta)k^*.
\]

Equation (107) describes the determination of \( k^* \) as a function of \( \sigma, n, \delta \) and \( f \). Figure 10.3 depicts equation (107) diagrammatically.
According to the Solow model, exogenous differences in saving rates ($\sigma$), 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.

**Differences in Saving Rates**

Using either Figure 13.2 or 13.3, 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 13.4 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 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.

Differences in Population Growth Rates
Again, using either Figure 13.2 or 13.3, 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 13.5 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.

![Figure 13.5: Per Capita Incomes and Population Growth Rates](image)

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.

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 13.3. 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).

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.

285
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.

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 13.6 displays the general equilibrium of this economy.
The triangular regions in Figure 13.6 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 13.7.
Point A in Figure 13.7 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 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?

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? 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.

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.\(^\text{110}\)

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.

A Simple Model

Let $z_t$ denote the stock of ‘knowledge capital’ available at date $t$, for $t = 1, 2, \ldots, \infty$. 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:

$$MRS(l, c_{t+1}(2)) = \frac{c_t(1) v'(l)}{\beta}.$$

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:

$$e + l = 1. \quad (108)$$

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:

$$z_{t+1} = (1 + e) z_t. \quad (109)$$
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 (109) and (108) to form a relationship that describes the trade off between current leisure and future consumption:

$$c_{t+1}(2) = (2 - l)z_t. \quad (110)$$

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,

$$\frac{c_{t+1}(2)v'(l)}{\beta} = z_t. \quad (111)$$

Conditions (111) and (110) are depicted in Figure 13.8.

**FIGURE 13.8**
Equilibrium Growth Rate

Notice that since $c_{t+1}(2) = z_{t+1}$, we can rewrite condition (111) as:

$$v'(l) = \beta \frac{z_t}{z_{t+1}} = \beta \left( \frac{1}{1 + e} \right),$$
or,

$$v'(1 - e^*) = \beta \left( \frac{1}{1 + e^*} \right).$$  \hspace{1cm} (112)

Equation (112) 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 (112) 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 (112) 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 13.9 displays the solution in (112) 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 13.9.

![Figure 13.9](image)

**FIGURE 13.9**

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 10.6, 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 \( MRS = \frac{c_{t+1}}{\beta c_t} \), 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:

\[
R^* = \frac{1}{\beta} (1 + e^*). \tag{113}
\]

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 \( z^A_1 > z^B_1 \), 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 \( z^A_1 \)). 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 14.1, 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).\(^{111}\)

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

\(^{111}\)Remember that since Figure 14.1 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.
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 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?