A Short Course in Introductory Macroeconomics

Not to be circulated

1. MEASURING THE MACROECONOMY I: GROSS DOMESTIC PRODUCT **4**

1.1. GDP AND NATIONAL INCOME ACCOUNTING 4
1.2. GDP AS A MEASURE OF STANDARD OF LIVING 6
1.3. PRODUCTIVITY 10
1.4. BUSINESS CYCLES 11

2. MEASURING THE MACROECONOMY II: UNEMPLOYMENT **13**

2.1. CLASSIFYING THE POPULATION 13
2.2. TYPES OF UNEMPLOYMENT 14
2.3. MINIMUM WAGE AND UNEMPLOYMENT 16
2.4. A THEORY OF THE LABOR MARKET 17
2.5. SEARCH THEORY: AN ALTERNATIVE THEORY OF THE LABOR MARKET (OPTIONAL) 21

3. MEASURING THE MACROECONOMY III: PRICES **22**

3.1. CONSUMER PRICE INDEX 22
3.2. INFLATION 23
3.3. BIASES IN THE CPI AND POLICY IMPLICATIONS 24

4. NOMINAL VARIABLES VS. REAL VARIABLES **26**

4.1. NOMINAL GDP VS. REAL GDP 26
4.2. GDP DEFLATOR – AN ALTERNATIVE MEASURE OF THE PRICE LEVEL 27
4.3. NOMINAL INTEREST RATES VERSUS REAL INTEREST RATES 28
4.4. DISTRIBUTIONAL CONSEQUENCES OF UNEXPECTED INFLATION 29

5. AGGREGATE DEMAND I: CONSUMPTION **31**

5.1. KEYNESIAN CONSUMPTION FUNCTION 32
5.2. PERMANENT INCOME HYPOTHESIS 33

6. AGGREGATE DEMAND II: INVESTMENT **36**

6.1. BENEFITS OF INVESTMENT 36
6.2. COSTS OF INVESTMENT 38
6.3. OPTIMAL INVESTMENT 39
6.4. MACROECONOMIC INVESTMENT DEMAND 41

7. AGGREGATE DEMAND III: DEMAND-SIDE EQUILIBRIUM **44**

7.1. AGGREGATE EXPENDITURE 44
7.2. INCOME-EXPENDITURE DIAGRAM 45
7.3. THE AGGREGATE DEMAND CURVE 46
7.4. THE KEYNESIAN MULTIPLIER 48
7.5. RECESSIONARY AND INFLATIONARY GAPS 50
7.6. SO HOW VALID IS THE NATIONAL INCOME IDENTITY? 50

8. AGGREGATE SUPPLY **52**

8.1. SHORT-RUN AGGREGATE SUPPLY 52
8.2. LONG-RUN AGGREGATE SUPPLY 55
8.3. RECONCILIATION OF LONG-RUN AND SHORT-RUN AGGREGATE SUPPLY? 56

9. GENERAL EQUILIBRIUM **58**

9.1. RECESSIONARY GAP 59
9.2. INFLATIONARY GAP 59
9.3. DID THE U.S. ENTER A "NEW ECONOMY" DURING THE 1990S? 60
9.4. GOVERNMENT STABILIZATION POLICY 61

10. FISCAL POLICY **62**

10.1. Revenue Sources and Expenditure Categories for the Government 62
10.2. Government Deficits and Government Debt 62
10.3. Fiscal Policy Effects on Aggregate Demand 63
10.4. Supply-Side Fiscal Policy 64
10.5. Effects of Fiscal Policy on Interest Rates and Investment 68
10.6. Limits on the Effectiveness of Fiscal Policy 72

11. MONEY AND THE MONEY MARKET... 74

11.1. The Role of Money 74
11.2. Measuring Money.. 74
11.3. Financial Intermediaries 75
11.4. The Money Market... 77

12. MONETARY POLICY.. 82

12.1. The Federal Reserve.. 82
12.2. Transmission Mechanism of Monetary Policy 85

13. PHILLIPS CURVE 89

13.1. Expectations Formation 90
13.2. Adaptive Expectations and the Clockwise Loop of Inflation Dynamics.......................... 91
13.3. Rational Expectations and the Vertical Long-Run Phillips Curve 92

14. INTERNATIONAL ECONOMICS I: INTERNATIONAL TRADE 93

14.1. Absolute Advantage and Comparative Advantage.. 93
14.2. Determination of World Prices 95
14.3. Distributional Consequences of Trade 96
14.4. Barriers to Trade 97

15. INTERNATIONAL ECONOMICS II: INTERNATIONAL FINANCE....................................... 99

15.1. Floating Exchange Rates.. 100
15.2. Currency Appreciation and Depreciation 102
15.3. Factors Affecting Exchange Rates... 103
15.4. Fixed Exchange Rates 104

16. INTERNATIONAL ECONOMICS III: OPEN ECONOMY MACROECONOMIC POLICY 110

16.1. Macroeconomic Effects of Currency Appreciation.. 110
16.2. Macroeconomic Effects of Currency Depreciation.. 110
16.3. Impact of Interest Rates on Capital Flows.. 111
16.4. Expansionary Fiscal Policy 112
16.5. Expansionary Monetary Policy 113
16.6. The “Twin Deficits” 114
16.7. The Asian Financial Crisis 116

17. ECONOMIC GROWTH 117

17.1. Neoclassical Growth Model (aka Solow Growth Model) 117
17.2. Endogenous Growth Theory (aka New Growth Theory) (OPTIONAL) 124

1. Measuring the Macroeconomy I: Gross Domestic Product

There are many important measures used to assess the performance of an economy at the macroeconomic level over time. Some of the most important measures that we will study are gross domestic product (GDP) and its components, the growth rate of GDP, the price level and the inflation rate, the unemployment rate, and the rate of technological progress.

1.1. GDP and National Income Accounting

GDP is the standard measure of economic activity for a country during a given time period. It is defined as the value of aggregate production of final goods and services in a country during a given time period. The notion of value that is used in the computation of GDP is market prices. That is, all quantities of goods and services are converted into dollar terms using market prices, and those dollar terms are summed up to yield (nominal) GDP.[1]

Because GDP is a measurement of activity during some time period, it is a *flow variable,* as opposed to a *stock variable,* which provides a measurement (of the capital stock of the economy, say) at a particular point in time. As a point of reference, U.S. nominal GDP for the calendar year 2000 was roughly $10 trillion. In the U.S., data on GDP are generally compiled and released on a quarterly (three-month) basis.

An important thing to keep in mind about the definition of GDP: GDP only includes those economic activities that are coordinated **through markets**. Thus, there are many "economic activities" that occur which do not get counted in GDP. Examples of these include activities in the household such as cooking or cleaning or child-rearing, as well as transactions between individuals that go unrecorded, such as a 15-year-old boy earning money from neighbors for shoveling driveways after a snowstorm. Thus GDP understates the level of economic activity.

There are three approaches to calculating GDP: the expenditure approach, the income/factor payments approach, and the production/value-added approach.

1.1.1. Expenditure Approach

The expenditure approach is the most straightforward approach to computing GDP and is often the most useful approach for macroeconomic analysis. The method relies on the basic national income identity:

$$GDP = C + I + G + NX,$$

where $NX = X - IM$ is net exports, defined as the difference between exports and imports. The other variables which appear in the national income identity are:

[1] The distinction between *nominal GDP* and *real GDP* will be made below.

- C: aggregate consumption
- I: aggregate investment; note that investment here does *not* refer to, for example, an individual purchasing a stock. This latter type of transaction gets categorized as savings, not investment. In macroeconomics, the term investment is reserved for purchases by firms of capital (machines, buildings, etc) and purchases by households of some durable goods (such as housing). We will study investment activity in more detail later.
- G: total government spending (includes both federal and local governments); does *not* include transfers payment by the government, such as Social Security or Medicare payments.

1.1.2. Income/Factor Payments Approach

This method of computing GDP asks the following question for all goods and services produced in the economy during a given time period: What incomes resulted from production of this particular good or service? The categories of income that result from production are wages, interest, rents, and profits. When summed over all goods and services produced, the resulting figure is the GDP. That is,

GDP = All wage payments + All interest payment + All rents + All economic profits.

If we adopt the model of perfect competition, in which there are zero rents and zero economic profits,[2] GDP is simply the sum of all wage payments and all interest payments, which can be represented as

$$GDP = w \cdot L + r \cdot K,$$

where w is the wage rate (set in a perfectly-competitive labor market), L is the total amount of labor supplied in the economy, r is the rental rate (interest rate) on capital (set in a perfectly-competitive capital market), and K is the total amount of capital available for production.

1.1.3. Production/Value-Added Approach

The **value added** by a firm is its revenue from selling a product minus the amount paid for goods and services purchased from other firms. For any firm, its value added **equals** wages paid + interest paid + rent paid + profits earned, because this sum is precisely the sales revenue of a firm minus its purchases from other firms – but this latter expression is the same as the GDP expression from section 5. Thus, GDP equals the sum of value added by all firms.

In this method, *every* purchase of a good or service is counted (rather than purchases of only final goods), but we do not count the entire selling price, only the portion that represents value added.

---

[2] Recall from basic microeconomics the distinction between economic profits and accounting profits, as well as the fact that in the long-run, perfectly competitive firms earn zero profits.

1.2. GDP as a Measure of Standard of Living

A standard measure of a country's "standard of living" is GDP per person (also known as per capita GDP). As mentioned above, GDP measures only those transactions coordinated through markets. In poorer countries (where "poorer" is measured by GDP), a higher percentage of economic activity occurs outside organized markets than in richer countries. Thus, in poorer countries, GDP understates economic activity by more than in richer countries. In addition, GDP does not take into account other non-market sources of utility, such as health, leisure time, political freedoms, and environmental qualities. Many studies have tried to incorporate these other factors into "more complete" measures of economic activity/well-being, but, not surprisingly, it is very difficult to quantify the (dollar) value of these non-market activities.

1.2.1. Economic Growth

Economic growth is measured by growth of per capita GDP over time. Conceptually, economic growth can be thought of as the expansion over time of the country's PPF for all goods and services.

The topic of economic growth is very important because of its long-term implications for the standard of living of a country's residents. There are large consequences of a sustained growth rate of 2% versus a growth rate of 4%, even though the difference sounds "small." The difference arises through the magic of compounding: consider two economies, one whose growth rate is 2% per year and other whose growth rate is 4% per year. At present, both economies enjoy the same standard of living (i.e., they have the same per capita GDP). After forty years, through compounding, the GDP in the slow-growing country will be 2.21 times as large as initially ($(1.02)^{40} = 2.21$), whereas GDP in the fast-growing country will be 4.80 times as large as initially ($(1.04)^{40} = 4.80$). This calculation shows that in less than two generations, GDP for the economy growing at 4% is more than double that for the economy growing at 2%, which is a large difference arising from a seemingly "small" difference in growth rates.

The study of economic growth is concerned with how an economy develops over long periods of time. As discussed above, although it can be criticized on several fronts, the main measure of economic standard of living used is real GDP per capita. Thus, economic growth will be measured as growth in real GDP per capita. For much of history, there was essentially zero economic growth. But over the past two to three centuries, economic growth has been positive, though growth rates still vary widely from one region of the world to another, and from one country to another.

There has been much debate about whether growth rates should imply "convergence" of real GDP per capita across countries. Evidence on this issue has been mixed. Many industrialized nations have indeed more or less converged to the same standard of living.

However, many developing countries seem to be "stuck" at far lower standards of living, though some evidence of "catch-up" has been witnessed in Asian economies.[3]

Economists generally believe that there are three preconditions for economic growth. These preconditions are:

1. Markets
2. Property rights
3. Monetary exchange

When considering the topic of economic growth, an important concept is that of **potential GDP.** Potential GDP is defined as the real GDP a country *could* produce if the labor force were fully employed.[45] The terminology is somewhat bad (because how can an economy go "above its potential," as it's seen to do empirically?). The sources of growth of potential GDP are: growth of the labor force, growth of the capital stock, and the rate of technological progress. We now turn to the subject of growth accounting, which allows us to separate economic growth into its three component sources. Note that growth accounting is **not** the study of economic growth – we will return to consider the topic of economic growth at the end of the course.

1.2.2. Growth Accounting using Aggregate Production Function

The commonly used abstract aggregate production for an economy, especially when considering issues of economic growth, is the general function $Y = AF\left(K, N\right)$, where Y is output, K is the capital stock, N is the size of the population,[6] and A is the level of technology.[7] Granted, this production function abstracts away from other inputs of production, such as land – however, it has the desirable features that it is relatively easy to work with mathematically, and it approximates actual production functions estimated using macroeconomic data in many countries not too badly.

Before proceeding, a further note about the "technology" A is in order. The terminology obviously evokes an image of machines, computers, etc. However, this is too restrictive an interpretation for our purposes. Because the only variables we are using in the specification $Y = AF(K, N)$ are output (Y), capital (K), and labor (N), the term A actually captures *all* other phenomena which impact the relationship between inputs and outputs. One of these phenomena is the advancement of knowledge. But A also captures features of an economy such as the level of government regulation, the form of

[3] See the graphs in the lecture slides.

[4] The terminology is somewhat bad (because how can an economy go "above its potential," as it's seen to do empirically?).

[5] More generally, potential GDP is the GDP a country can produce when *all* factors of production are fully employed.

[6] Or you can think of it as the size of the labor force. In the long-run, however, L and N will be strongly positively correlated, so to the extent that is true, it does not matter which variable we place in the production function.

[7] Also known as total factor productivity, or, as mentioned below, the Solow residual.

government, the political process, cultural phenomena, and other such related matters. For example, in two countries that are identical in every way except for the level of government interference in government – in one country, the government meddles in business affairs while in the other laissez-faire dominates – the country with less government interference would generally be expected to have higher output. Thus, K and N in the two economies would be the same, while Y would be higher in the country with less government interference in business. Provided the two countries have identical production functions $F(\bullet)$, we would say that the country with less government interference in business has a higher value of A, a higher level of "technology."

The method of **Solow growth accounting,** named for the developer of the method Robert Solow, attempts to decompose the total growth rate of output into the growth rates of the inputs and the growth rate of technology. In order to do this for an economy, one must assume a particular functional form for the aggregate production function. A commonly used form is the **Cobb-Douglas production function**

$$Y = AK^{1/3}N^{2/3}$$

The exponents on K and N above are the values found empirically to be true for the aggregate production function in the U.S. A more general approach is to simply express these exponents in abstract terms, in which case the production function would be written as $Y = AK^{1-\alpha}N^{\alpha}$, where the fraction α is labor's share of total production and $(1-\alpha)$ is capital's share of total production. In other countries, these exponents can and generally do have values different from those for the U.S. – the exponents simply represent the relative intensity of use of the respective input factor in production. For the following discussion, the empirical values for the U.S. are used – as a test of your understanding, you should re-derive the calculations that follow using the variables α and $(1-\alpha)$, rather than the specific values of 2/3 and 1/3, respectively.

Two features of the Cobb-Douglas production function that conform to empirical evidence are that it:

1. Displays **constant returns to scale.** Simply put, this feature says that if **all** inputs are scaled by a factor k, then total output is scaled by the factor k.
2. Displays **diminishing marginal product.** This is the feature that when all but one factor of production is held constant, the total output generated by increasing the remaining input increases at a decreasing rate. That is, when viewed as a function of only one of the inputs, the graph of output is as follows:

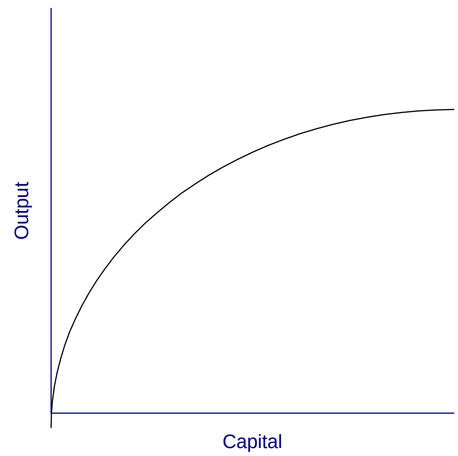

Figure 1

A practical issue is the issue of how to *measure* changes in technology – that is, how to measure changes in the value A. Everyone would agree that technology has improved over time, but when one begins to think about it, it is not at all clear how to accurately and meaningfully measure what technology is or how it changes over time.

The technique that Solow growth accounting uses is to measure technology change as the "unexplained" part of output growth. That is, it is fairly easy to obtain data about output, population size, and capital stock for an economy. Using the assumed production function then allows an estimation of the rate of technical progress – the rate of technical progress is taken to be simply that portion of output growth not accounted for by growth in the inputs.

Using the production function $Y = AK^{1/3}N^{2/3}$ for the U.S., this decomposition[8] yields

$$\%\Delta Y = \%\Delta A + \frac{1}{3}\%\Delta K + \frac{2}{3}\%\Delta N$$

or, after rearranging to isolate the technical progress term,

$$\%\Delta A = \%\Delta Y - \frac{1}{3}\%\Delta K - \frac{2}{3}\%\Delta N$$

To summarize, the growth accounting method decomposes total economic growth of output into growth of inputs and growth of technology. Because it is very difficult to measure "technology" directly, but capital and labor are fairly easy to measure, the rate of technological growth is taken to be that amount "left over" after the growth rates of the

[8] The following decomposition is purely a mathematical one. For any two variables x and y, a mathematical fact (one that we will not prove in this class) is that the percentage change in their product xy equals the sum of the percentage changes in each variable separately.

inputs have been taken into account. The term A is often called the **Solow residual** because it is measured using Solow growth accounting.

One recent practical application of growth accounting appeared in an article written by the noted economist Paul Krugman[9] before the Asian currency crisis in the late 1990's. His estimates of the rates of technical progress in Asian economies showed that technology was growing very slowly in that region – even more slowly than in the U.S., despite that fact that Asia was growing much more rapidly than the U.S. at the time. His study suggested that a very large fraction of the rapid growth could be explained by the large increases in the capital stock and the hours worked in the economy. This finding suggested that the Asian economies were not "a miracle" – rather they were flying high because people and machines were stretched to and beyond their capacity. The ensuing crises in Asian economies perhaps vindicates this view.

1.3. Productivity

Intimately related to issues of economic growth is the productivity of a country's resources, the most important of which is labor productivity. Labor productivity is defined as output per hour worked in the economy.

In the 1970's the U.S. experienced what has come to be called a great productivity slowdown. Some reasons commonly offered for the slowdown of the 1970's, along with arguments against them, are as follows:

1. Lagging investment (savings)
    - It is true that savings in the U.S. declined in the 1980's, BUT firms' investment did NOT decline, because the U.S. borrowed from abroad
2. High energy prices
    - Oil shocks in the 1970's did coincide with recessions in the U.S.
    - BUT – oil prices dropped in the 1980's and 1990's and productivity growth did not seem to pick up
3. Deteriorating workforce skills
    - U.S. SAT scores dropped in the 1980's
    - But as measured by the "college premium" (defined as average wage of college graduate/average wage of high school graduate), the quantity of "skilled" workers increased during the 1980's and 1990's!
4. Foreign Competition
    - Other nations, both developed and developing, began to compete more aggressively with the U.S. in many markets.
    - But ALL industrialized nations experienced productivity slowdowns essentially simultaneously
5. Adoption of new technology

[9] For interested readers, Krugman writes a very readable and often very insightful column twice-weekly on economic issues in the op-ed section of *The New York Times.*

- It takes time for workers and firms to learn how to use all the new information technology (Web, email, etc) efficiently
- It is too soon to tell if this explanation will be a good one or not

6. Mis-measurement of real GDP
   - This explanation may prove to be the most promising, especially because the production of services, which constitutes an ever-growing share of total GDP, is particularly hard to measure.

1.4. Business Cycles

While the topic of economic growth is concerned with changes in GDP over very long periods of time, it is an economic fact of life that GDP changes occur over much shorter time-horizons as well. The periodic ups and downs of economic activity are termed the "business cycle." Figure 2 shows the periodic movement of real GDP for the U.S. since 1959.

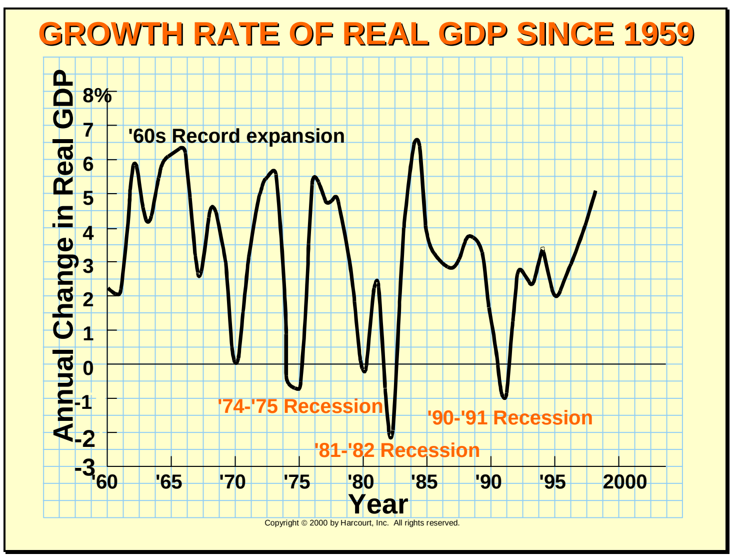

Figure 2

The movement from a peak to a trough is called a recession, and the movement from a trough to a peak is called an expansion. In the U.S., the average recession has lasted 11 months, and the trough has, on average, has been 2.5% lower than the real GDP of the previous peak.

Just as the study of long-term economic growth is one major branch of macroeconomics, so too is the study of the short-term fluctuations associated with the business cycle. Indeed, we will see that many of the policy prescriptions that macroeconomics offers, such as taxing and spending by Congress and control of the money supply by the Federal Reserve, are inherently short-term solutions to current economic conditions, rather than vehicles to promote long-term economic growth.

2. Measuring the Macroeconomy II: Unemployment

The study of unemployment is one major focus of macroeconomics. Unemployment interacts with the rest of the macroeconomy in often predictable (and sometimes unpredictable) ways. As such, it is important to understand some of the most important definitions and concepts used in the field of labor economics.

2.1. Classifying the Population

In order to measure the "unemployment rate," the Bureau of Labor Statistics (BLS) categorizes all people in the economy into two broad categories: inside the labor force and outside the labor force. In practice, the BLS surveys a representative cross-section of the country's population about their labor force status every month. From these surveys, the BLS extrapolates the results to the entire population. Individuals who are inside the labor force are further sub-categorized into two sub-groups: employed and unemployed. It is thus most useful to think of all individuals in an economy as belonging to one of three "pools" of individuals: employed, unemployed, and outside the labor force, as illustrated in Figure 3.

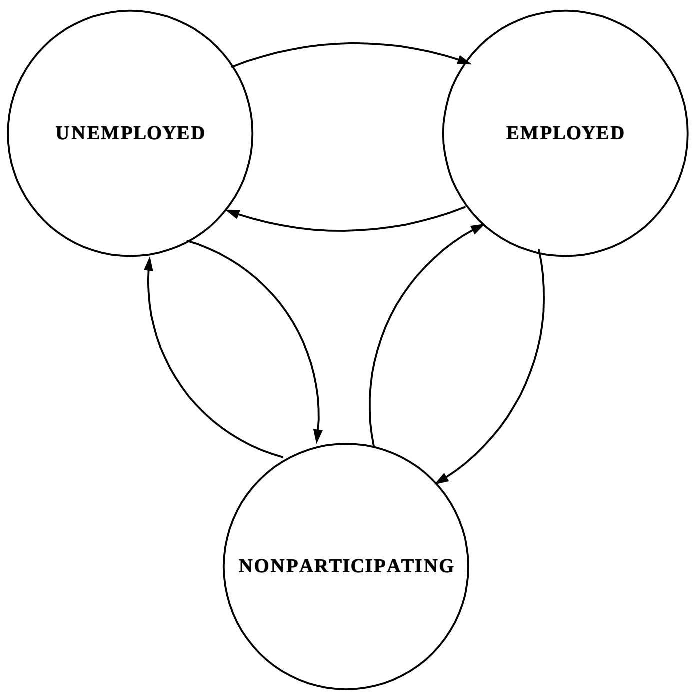

Figure 3

The figure shows that there exist flows in both directions between any two of the labor market pools. The unemployment rate is defined as the ratio of unemployed persons to the size of the labor force. That is,

$$UE\ rate = \frac{Number\ of\ unemployed\ individuals}{Number\ of\ individuals\ in\ the\ labor\ force}$$
$$= \frac{Number\ of\ unemployed\ individuals}{Number\ of\ unemployed\ individuals + Number\ of\ employed\ individuals}.$$

Notice that individuals who are outside the labor force, and are thus not working, are **not** counted as unemployed. This is of course simply a matter of definition, but there are good reasons for doing so. Individuals are classified as outside the labor force if they fail to meet at least one of the following criteria:

- Is available for work
- Is willing to work
- Has made an effort to find work in the previous work

Notice especially the last item listed above. An individual who is not currently working and has not made an effort recently to find work is considered to be outside the labor force. It is possible that an individual desires a job but has not made an effort recently because he believes it unlikely that he would find a job even if he did search. Such an individual is called a **discouraged worker.** Also note that workers looking for better jobs and part-time workers are counted as employed, not as unemployed. One phenomenon which has received media attention lately is "underemployment," which is a term meant to describe individuals who are not working as much, and thus not earning as much, as they would like.[10] The official labor statistics, however, simply count these individuals as employed.

The basic idea behind not counting persons outside the labor force in the unemployment rate is that such persons are, for one reason or another, not currently part of the normal productive resources of the economy. As such, they should not be considered when accounting for the used and unused resources of the economy.

From the definition of the unemployment rate given above, it should be clear that if an unemployed individual drops out of the labor force, the unemployment rate decreases, **even though the number of employed individuals has not decreased.** Thus, movements in the unemployment rate over time (especially short periods of time) do not necessarily indicate anything about how many people are employed in the economy.

2.2. Types of Unemployment

There are three categories of unemployment: frictional unemployment, structural unemployment, and cyclical unemployment.

[10] Often associated with the term "underemployment" in the popular media has been the term "living wage," which is meant to highlight the notion that many workers who work full-time and earn the minimum wage nonetheless do not earn enough to meet their basic expenses.

2.2.1. Frictional unemployment

Frictionally unemployed individuals are those individuals who are temporarily between jobs. For example, a worker may quit his current job with a plan to then search for a new job that he presumes he will enjoy more and the belief that he will indeed find one. This individual is thus temporarily unemployed.

2.2.2. Structural unemployment

Structural unemployment occurs when workers are dislocated from their jobs because of technological advances and/or international competition. The individuals who lose their jobs in such a manner are obviously adversely affected by such advances and competition. However, from a longer-term macroeconomic perspective, technological advances, which lead to structural unemployment, are the foundation of long-term economic growth. International competition that causes labor market dislocations also has long-term benefits for the entire economy because a country is able to leverage its comparative advantage to a greater degree the more open to international trade it is.[11]

2.2.3. Cyclical unemployment

Cyclical unemployment is generated by the periodic ups and downs of the business cycle. During periods of macroeconomic slowdown (i.e., recessions), unemployment generally rises because firms tend to cut back on their numbers of workers. During periods of economic expansion, unemployment generally falls as firms try to take advantage of the "good times" by hiring more workers to produce more output. Indeed, the experience of the United States since 1995 largely bears out these correlations. From 1995 through 2000, a time of great expansion for the U.S. economy, the unemployment rate fell. Since early 2001, however, the unemployment rate has been rising as economic activity has been slowing down.[12]

2.2.4. Natural rate of unemployment

Both frictional and structural unemployment are viewed as generally "good" for the economy. Frictionally unemployed workers are, for the most part, simply moving on to

---

[11] Recall from introductory microeconomics that an economy is able to reach a point beyond its production possibilities frontier (PPF) by trading with another country as long as the slopes of their PPFs are not identical. In other words, a country can reach a consumption possibilities frontier (CPF) outside its PPF by trading internationally, thus increasing the welfare of the economy.

[12] The National Bureau of Economic Research (NBER) declared in November 2001 that the U.S. economy had entered recession in March 2001. Note that a recession is not strictly defined as two consecutive quarters of negative GDP growth, as is often cited in the popular media. As of June 2003, the NBER still had not declared the "official" end of this recession, despite the opinion of many observers that it had ended by the end of 2001..

better and more rewarding jobs. This process creates an allocation of workers to jobs which is overall more productive for the economy. And while workers who lose their jobs due to technological progress and/or international competition suffer, especially if there is little or no social safety net to assist them, the overall macroeconomy benefits in the long run due to such structural reallocations. These two types of unemployment, thus, are called **natural unemployment,** and, accordingly, the natural rate of unemployment is defined as the sum of the frictional unemployment rate and the structural unemployment rate.

When the only unemployment in an economy is of the frictional and structural types, the economy is said to be at **full employment.** Note that the concept of full employment does **not** mean that every person in the economy has a job. Economists believe this is a sensible notion of "full employment" because it allows the economy to be flexible enough to respond to changes in shocks that affect the economy. If instead the notion of "full employment" was intended to capture the case that literally everybody in the economy who wanted a job had one, then it would be difficult, for example, for a new sector of the economy which all of a sudden had tremendous growth prospects to hire enough workers to grow to its full potential. Many economists believe that were it not for natural unemployment, the tremendous growth of the dot-com industries in the 1990's could not have happened. Specifically, if everyone who wanted a job had one, there would have been fewer individuals who would have been willing and/or able to move into the technology sector, which quite possibly would have meant that (the recent dot-com-bust notwithstanding) technology would not have advanced as rapidly as it has over the past half-decade.[13]

2.3. Minimum Wage and Unemployment

One common theory for the existence of unemployment employs a simple supply-demand analysis. As in any other competitive market, the price of labor (which **is** what is called the "wage") is determined as that price such that the supply of labor equals the demand for labor. We will explore this theory below.

Many countries, including the U.S., have a minimum wage law. Simple microeconomic theory tells us that if this minimum wage is set above the free-market equilibrium wage, then the supply of labor (i.e., the number of people who are willing, able, and searching for work) will be larger than the demand for labor (i.e., the number of people who actually do work). This excess supply of labor is unemployment. Thus, one line of reasoning holds that the abolition of minimum wage laws would greatly **reduce** the unemployment rate, because their abolition will allow labor markets to come into equilibrium, in which the supply of labor equals the demand for labor, which, by definition, would mean that there is no unemployment.

[13] It is also true that the booming economy of the late 1990's also drew more workers than ever before into the labor force. But for the purposes of this discussion, we ignore this aspect.

2.4. A Theory of the Labor Market

Because unemployment is one of the major topics in macroeconomics, it is important to understand how the labor market functions. In particular, the labor market is a *market*, like any other market in many respects, that you are familiar with from introductory microeconomics. Thus, there are two sides to the labor market – the demand for labor, and the supply of labor.

2.4.1. Labor Demand

The demand for labor is determined through the profit maximization problem of firms. Recall that profit is defined as total revenue minus total cost – symbolically, as $P \cdot Y - w \cdot L$, where P is the price of the output good, Y is the output of the firms, w is the wage rate (meaning hourly wage), and L is the amount of labor (you can think of it as total hours of work) hired by firms.

Notice we are assuming that the only input needed for production is labor. In general, of course, this is not true – in general, production depends on many other factors, such as capital, technology, and other inputs, as mentioned previously. But for the moment let's make the simplifying assumption that the only input needed is labor input.

Thus, the decision of firms involves choosing how much labor L to hire. So it remains to determine how firms will make their choice. Firms will hire the amount of labor such that the wage rate w = the marginal revenue product (recall this from introductory microeconomics). That is, firms will hire an amount of labor L such that

$$P \cdot MP_L = w \tag{0.1}$$

This condition makes sense because if $P \cdot MP_L > W$ were true, then it would be worthwhile for firms to hire additional labor because the extra amount they would generate in revenue ($P \cdot MP_L$) exceeds the extra cost w of hiring one more unit of labor. The argument works in the reverse direction for the case if $P \cdot MP_L < W$, and firms would want to hire less labor. Thus, condition (0.1) defines the amount of labor that firms will demand. (Crucial to this preceding argument is the fact that the production function $f(L)$ that we are assuming displays diminishing returns – which is the reason that MP_L decreases as L increases.)

Rearranging (0.1), we arrive at

$$MP_L = \frac{W}{P} \qquad (0.2).$$

The quantity on the right-hand side of (0.2) is the real wage – that is, the amount of goods that the money-denominated wage W can buy. Each side of expression (0.2) can be plotted against labor (L) to generate the diagram in Figure 4:

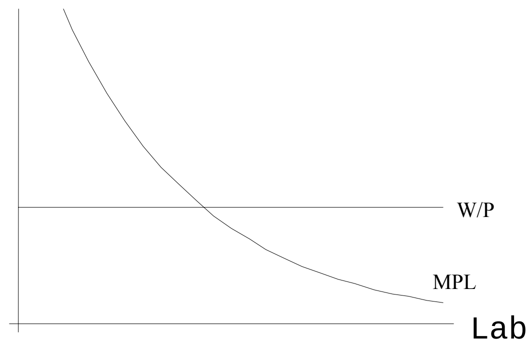

Figure 4

The curve labeled MP_L *is* the labor demand curve – thus, we will denote it as L^D below.

Factors that shift the labor demand curve include improvements in technology and an increase in the quantity of capital, which (all else being equal) raises the productivity of workers.

2.4.2. Labor Supply

The labor supply decision of households (consumers) comes from their utility-maximization problem. Here, we assume that households derive utility from both consumption and leisure (which is the number of hours spent *not* working) – that is,

$$utility = u(c, leisure) .$$

Naturally, because the total amount of time in, say, one week is fixed for every household, the relationship between time spent working and time spent enjoying leisure is

$$labor\,\sup ply = total\,available\,time - leisure .$$

Again from introductory microeconomics, the optimality condition, which determines how much leisure a household will desire and therefore how much work a household should supply to economic markets, is

$$\frac{MU\ of\ leisure}{MU\ of\ consumption} = \frac{price\,of\ leisure}{price\,of\ consumption} ,$$

where MU denotes marginal utility. In macroeconomics, the price of consumption is simply the price level P (as measured by, say, the CPI or the GDP deflator – we will consider the notion of the price level of the economy more carefully below). The price of leisure is the wage that could have been earned by working instead – that is, the price of leisure is the *opportunity cost* of enjoying that leisure. Thus, the above expression can be written as

$$\frac{MU\ of\ leisure}{MU\ of\ consumption} = \frac{W}{P} \tag{0.3}$$

It should seem reasonable that the higher the real wage, W/P, the more labor a household will be willing to supply. Thus, the labor supply function can be written as

$$labor\ \text{sup}\ plied = L^s(W/P) \tag{0.4}$$

To be more realistic, the amount of labor a household supplies is probably a function of more variables than simply the real wage. In particular, we may posit a labor supply function such as

$$labor\ \text{sup}\ ply = L^s(W/P, wealth, r)$$

in which labor supply depends positively on the real wage, negatively on wealth, and positively on the real interest rate r.[14] However, for simplicity, we will use the labor supply function given in (0.4). Plotting (0.4), we get Figure 5:

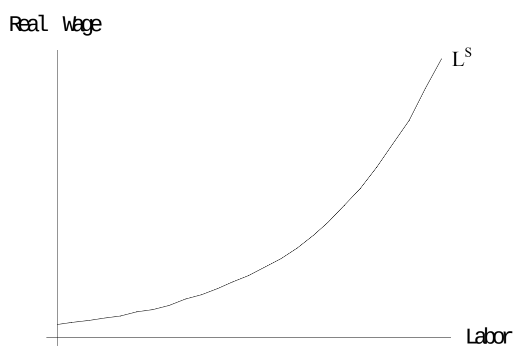

Figure 5

2.4.3. Labor Market Equilibrium

[14] We will study interest rates more carefully below – for now, simply think of the relationship as the higher is the interest rate, the more incentive there is to work and save in the present.

We can now describe equilibrium in the labor market. Notice that both labor demand and labor supply depend, as we have described them, only on the real wage W/P. Superimposing the above two figures, we get Figure 6:

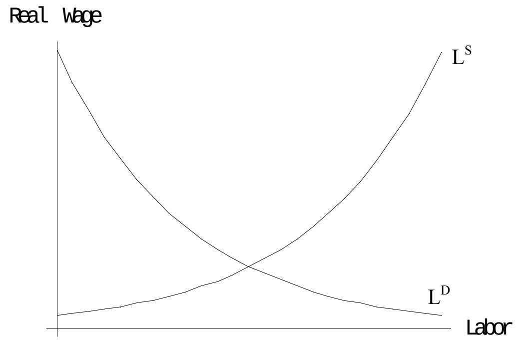

Figure 6

The most useful (and correct) way to think about this diagram is simply as the supply and demand curves in any other market, such as the kinds you are familiar with from introductory microeconomics.

Labor market equilibrium thus occurs at the point where the two curves intersect – and the equilibrium is described by a quantity of labor L* and a real wage $(W/P)^*$.

2.4.4. Labor Market Rigidities Cause Unemployment

An aspect of reality that the above simple exercise fails to capture is the presence of unemployment. According the standard competitive model (again, as in introductory microeconomics), the price level of the labor market (that is, the real wage) will adjust so that labor supply and labor demand exactly equal each other. But the implication of this is that there is no unemployment! In other words, each agent that is looking for a job can find one, at the real wage $(W/P)^*$. Thus, some reconciliation of this standard microeconomic story must be made with the empirical fact that unemployment does exist in the real world.

The mechanism used in order to reconcile the theory with the facts is that prices (wages) in labor markets are usually set in advance, often through bargaining and contracts. For example, the (real) wage that an autoworker earns today may have actually been determined by a contract that took effect a year ago. This contract, and the nominal wage W, was presumably set with some expectation of inflation in mind. If actual inflation turns out to be different from expected inflation, then the *actual* real wage earned today by the autoworker will be different than the one that should clear the labor market. In particular, if the actual real wage is higher than the real wage that would clear the labor market, labor supply will exceed labor demand (refer to the figure above) – which is the definition of unemployment.

Alternatively, the presence of a minimum wage law that prescribes a minimum wage higher than the market-clearing wage will cause labor supply to exceed labor demand – again, unemployment.[15]

Both of these phenomena – minimum wage laws and the presence of long-term labor contracts – are sources of **rigidities** in the labor market.

2.5. Search Theory: An Alternative Theory of the Labor Market (optional)

An alternative way to theoretically model the labor market that does not resort to the ad-hoc mechanism of introducing labor market rigidities in order to generate unemployment considers the actual decision faced by an individual when deciding whether or not to accept a job offer.

The concept of an individual's **reservation wage** describes the notion that any particular wage offer received by an individual falls either above or below some cutoff value that the individual has in mind. If a wage offer is above an individual's reservation wage, the individual will accept the offer and begin working – if the offer is below the reservation wage, the individual will reject the offer and keep searching for another wage (job) offer. An important point is the *each individual has his or her own reservation wage* – there is no reason why all individuals in the economy should have the same reservation wage. This approach to modeling the labor market does seem to have appealing properties, as verified simply through personal experience.

Thus, "unemployment" is a natural feature of this model, rather than one that is ad-hoc. Because some workers who are searching for jobs encounter jobs that they simply do not want (i.e., wages which are below their reservation wages), they *choose* to remain unemployed. An important lesson of this model, then, is that it may be *optimal* for some workers to remain unemployed.

[15] Convince yourself of this point using Figure 6.

3. Measuring the Macroeconomy III: Prices

Another major issue in macroeconomic analysis is how prices, at some aggregate or average level, evolve over time. In microeconomic analysis, thinking about the price of a single good is usually not problematic – the price is simply measured as the amount of some other item (usually money) that must be given up for one unit of that good. In macroeconomics, however, there is no single good to which consumption refers. That is, the C which represents consumption in the accounting identity $Y = C + I + G + NX$ does not refer to any particular good, but rather to some theoretical aggregate good.[16] Thus, the aggregate price level in the economy must be the price of this theoretical aggregate good. As such, we will speak of the price level of the economy, and denote it in theoretical discussions by P. But because this price is not attached to any good which actually exists in the economy, we will be often be more concerned with changes in the price level over time. The rate of change of the price level in an economy over time is what is known as inflation, which is a concept familiar from experience. We now turn to a more detailed examination of how the price level in an economy is measured.

3.1. Consumer Price Index

The Consumer Price Index (CPI), compiled monthly by the U.S. Bureau of Labor Statistics, is one of the most-widely watched measures of the price level. The rate of change of the CPI is the most common measure of inflation reported in the media. The CPI, as its name designates, is an index of the price level. It represents neither the price of any one particular good or service nor the prices of all goods or services. Rather, the most commonly followed component of the CPI is a weighted average of the prices of a basket (collection) of goods that are meant to be representative of the purchases of a typical urban family.[17] This market basket, which currently contains 400 items, is revised periodically to reflect changes in spending habits.[18]

BLS workers conduct monthly surveys of the prices of the goods in the basket in urban centers throughout the U.S. In order to measure changes in prices, the quantities of the goods purchased over time must somehow be held constant[19] – which is exactly what using the fixed market basket accomplishes.

[16] A note is in order here. Recall that every item, including consumption, in the national income identity is actually measured in dollars, not in terms of goods. But recall that this was needed simply to convert all of the many goods and services into a common unit of measurement. Conceptually, consumption (as well as all of the items in the income identity) should indeed be thought of as real goods and services.
[17] This most commonly-watched version of the CPI is referred to as the "CPI – all goods."
[18] The marker basket used to be updated every 10 years. However, with ever-more-rapidly changing consumer tastes and available products, the BLS recently announced that it would update the market baskets more frequently.
[19] Otherwise it would be impossible to isolate changes in prices from changes in quantities, because consumption and GDP are ultimately measured in terms of the *sum of products* of prices and quantities of individual goods.

Once the prices at some particular point in time of all the goods in the market basket have been obtained, it is a simple matter to compute the total cost of the market basket at that point in time – it is, of course, simply the sum of the products of all the prices with their respective quantities. The number computed in this way could serve as a measure of the price level. However, because the cost of this market basket would usually be some cumbersome number (in the thousands or tens of thousands of dollars, perhaps), this cost is converted into a more convenient measure using some arbitrarily chosen period of time as the base period. Specifically, the expression that yields the reported value of the CPI for some time period t is given by

$$CPI_t = 100 \cdot \frac{\text{cost of market basket in period t}}{\text{cost of market basket in base period}}.$$

By computing the CPI in this way, the measure of the price level is a pure number, i.e., one that has no units attached to it.[20] This is another reason why the CPI is an *index number*, and in fact this characterization is often taken to be the defining feature of an index number.

As we will see soon when we study the GDP deflator, the CPI is not the only empirical measure of the price level of the economy.

3.2. Inflation

Inflation is a general rise in the price level over time. Formally, it is the rate of change of the price level during some specified period of time. In any period t, the inflation rate relative to period $t-1$ is defined as

$$\pi_t = \frac{P_t - P_{t-1}}{P_{t-1}},$$

where π denotes the inflation rate.[21] As a matter of terminology, **deflation** (negative inflation) occurs when $\pi < 0$, and **disinflation** occurs when π decreases over time (but is still positive at every point in time). For example, if in four consecutive years, inflation was 20%, 15%, 10%, and 5%, we say that disinflation is occurring – even though the price level increased in each of the four years.

In practice, the CPI is usually used to compute and report the inflation rate by the popular media, which once again emphasizes that the actual number reported as the CPI is of little consequence – the main purpose of computing a price index is to compute its rate of change.

3.2.1. Costs of Inflation

[20] Because, notice, the units of dollars in the numerator and denominator of the second term on the right hand side of the previous expression cancel with each other, leaving a unitless number, otherwise known as a pure number.
[21] Not to be confused with profits, which is what π typically represents in microeconomics. The usage is almost always clear from the context.

There is much debate about whether inflation is costly for society. Recall from microeconomics that prices are the signals by which consumers and firms learn about the relative scarcity of different goods. Price adjustments are the mechanism by which perfectly competitive markets reach equilibrium, in which suppliers' decisions are consistent with demanders' decisions. For inflation to be costly to society, then, where "costly" should be interpreted to mean a decrease in overall economic efficiency, inflation must be unpredictable. If inflation is unpredictable, then the price signals sent by markets are the incorrect ones, reducing economic efficiency. However, to the extent that inflation is predictable, it is often argued that there seems to be no rational reason why consumers and firms cannot simply take expected inflation into account when making decisions. Nonetheless, it is the unpredictability of inflation that makes it one of the most important issues in macroeconomics.

Even if inflation is predictable, however, there are scenarios in which inflation imposes a cost on society. This typically occurs during periods of **hyperinflation**, which is loosely defined as periods when the price level rises by hundreds or thousands of percent per year. Two notable hyperinflations in the 20th century occurred in Germany in the 1930's and in Argentina in the 1980's. When a country is experiencing hyperinflation, consumers will rationally try to unload their money as quickly as possible to buy the goods and services they desire. Imagine a spectacular hyperinflation in which the price of bread doubles in one day. In such a scenario, workers may demand that employers pay them daily or even multiple times per day, and, once they have been paid, rush to stores to purchase bread in an attempt to avoid the expected increase in the price. Such activity obviously reduces the amount of output workers will produce (because they are constantly rushing to the markets to purchase goods and services), and the cost of the loss in output is termed **shoeleather costs.** Another type of cost is borne by society in periods of rapid of inflation, this one by firms. In the above example, the sellers of bread will have to print new signs posting the price of bread daily – which is an unproductive use of materials and time. This latter type of cost is called **menu costs.**

The consensus view among the economics profession, then, is that moderate rates of (mostly) expected inflation do not impose a large welfare cost on society. However, very rapid inflation does reduce welfare.

3.3. Biases in the CPI and Policy Implications

In the mid-1990's, an important report concerning the CPI was issued by the Boskin commission, a group appointed by the Senate Finance Committee.[22] Its main finding was that the CPI overstates the inflation rate by 1.1% per year. The commission cited three major reasons why the CPI overstates inflation. The three sources of bias in the CPI are:

1. **New goods bias** – The CPI basket is updated once every 10 years, which leads to newer (and often cheaper) goods being omitted, even though

[22] The full report is available at http://www.ssa.gov/history/reports/boskinrpt.html.

consumers may in fact have stopped purchasing some of the goods in the official basket.

2. **Quality change bias –** If a good or service simply becomes a better product over time, then price increases may reflect the improved quality, rather than inflation. For example, computers get faster and better very frequently, but "buying a computer" means the same thing from period to period. That is, the CPI's market basket may simply treat computers as a homogenous product, regardless of quality improvements over time. As a result, even if consumers are actually spending more for "a computer," they may in fact be getting a much different (and better) product, a confounding effect that a good measure of inflation would be able to isolate.
3. **Commodity substitution bias –** The CPI uses a fixed basket, without taking account of the fact that as the price of a good rises, consumers may substitute away from that good. For example, when the price of Pepsi rises, many people may switch to Coca-Cola. However, if Pepsi is included in the market basket and Coca-Cola is not, the CPI will overstate the rate of change of consumers' actual expenditures.

One important policy consideration surrounding the computation and biases of the CPI concerns payments to recipients of social security benefits and veterans' benefits. These benefit payments are adjusted periodically to keep up with changes in the cost of living, and the CPI is used to compute the change in the cost of living. Thus, the upward biases in the CPI tend to **increase** government outlays for social security. One implication of computing the CPI in such a way as to take account of the findings of the Boskin commission is that benefit recipients would receive smaller periodic increases in their transfer payments.

4. Nominal Variables vs. Real Variables

Thus far, we have been a bit abstract in referring to GDP. In our definition of GDP, we stated that it was the "value" of all goods and services produced during a given time period, and we implicitly have been supposing that it is money values that the definition requires. In fact, this is not so. Money units are ultimately simply "units" that society finds convenient when measuring economic value. It is a measuring unit just like, for example, feet and pounds are measuring units. There is nothing fundamentally "correct" about measuring length in feet (or meters) or weight in pounds (or kilograms). Analogously, there is nothing fundamentally "correct" about measuring economic value in money units.

In this section, we explore the distinction between **nominal economic variables** and **real economic variables.** The two most important instances of this general distinction concerns GDP and interest rates.

4.1. Nominal GDP vs. Real GDP

If we sum the multiplicative products of money prices and quantities of all goods and services produced in an economy in a given time period, we arrive at **nominal GDP.** For example, suppose that there are three products produced in an economy, good 1, good 2, and good 3. The dollar price of good 1 in the year 2001 is P_1^{2001}, the dollar price of good 2 in the year 2001 is P_2^{2001}, and the dollar price of good 3 in the year 2001 is P_3^{2001}. Further, suppose that the same three goods are also the only goods produced in the year 2002, and their respective dollar prices in 2002 are P_1^{2002}, P_2^{2002}, and P_3^{2002}. We would compute nominal GDP in the year 2001 as

$$\text{Nom GDP in 2001} = P_1^{2001}q_1^{2001} + P_2^{2001}q_2^{2001} + P_3^{2001}q_3^{2001},$$

where the q terms represent the quantities of each good produced. Similarly, we would compute nominal GDP in the year 2002 as

$$\text{Nom GDP in 2002} = P_1^{2002}q_1^{2002} + P_2^{2002}q_2^{2002} + P_3^{2002}q_3^{2002}.$$

Comparing these last two expressions, it should be clear that there are two main ways that nominal GDP can change from one year to the next: at one extreme, prices may change but quantities may remain constant, and the other extreme, prices may remain constant, but quantities may change.

The main purpose of constructing a measure like GDP is usually to measure quantities. As such, any change in prices between 2001 and 2002 in the above example confounds our attempt to do so. However, suppose we compute the following

$$P_1^{2001}q_1^{2002} + P_2^{2001}q_2^{2002} + P_3^{2001}q_3^{2002},$$

in which we are using year 2001 prices in conjunction with year 2002 quantities. In this way, the measure we construct captures only the effects of changes in actual output, not some combination of changes in prices and changes in output. We call the last expression **real GDP** because it is computed using constant prices (in this case, year 2001 prices).

We now describe the general conceptual difference between nominal GDP and real GDP. Nominal GDP measures the value of a nation's output during a given time period using that time period's money prices. Real GDP measures the value of a nation's output during a given time period using the money prices of some base (or reference period). In the example above, the base period was 2001.

4.2. GDP Deflator – an alternative measure of the price level

Now that we have made the distinction between nominal GDP and real GDP, we introduce an alternative measure of the price level, the **GDP deflator**. The GDP deflator is defined in the following way:

$$\text{GDP Deflator} = \frac{\text{nominal GDP}}{\text{real GDP}} \cdot 100.$$

It is often argued that the GDP deflator is a better measure of the price level than the CPI is because the CPI uses a fixed basket of goods, whereas the GDP deflator uses *all* goods and services produced in the economy. As Figure 7 shows, the CPI and GDP deflator have tracked each other fairly closely in the U.S. for the past 50 years, although the GDP deflator shows less volatility, presumably because of the broader basket of goods that it uses.[23]

[23] As a quick test of your understanding: what must be the value of the GDP deflator for the base period (whenever that may happen to be)?

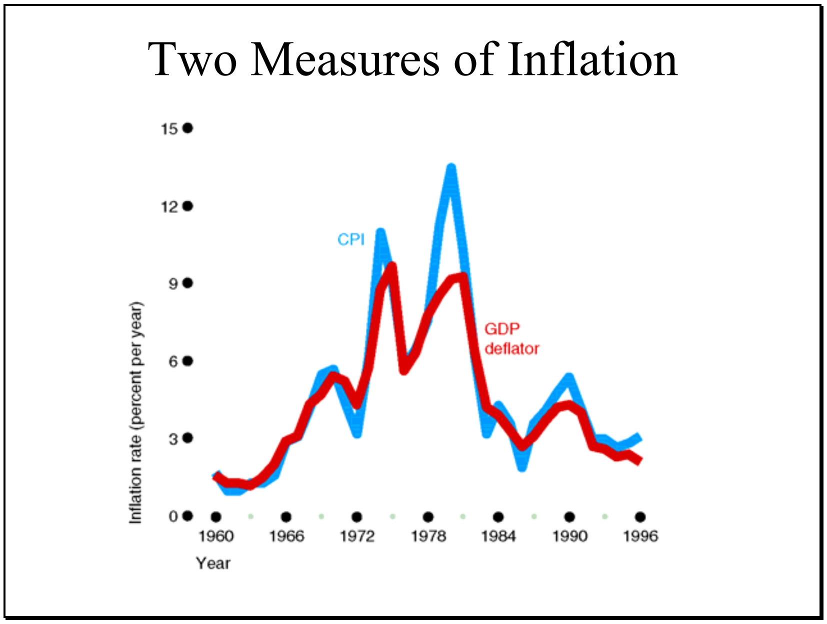

Figure 7

4.3. Nominal Interest Rates versus Real Interest Rates

There is another important nominal vs. real distinction to introduce, the distinction between nominal and real interest rates. Most interest rates that are observed in everyday economic life are nominal interest rates. For example, a savings account at a bank that features a three percent annual interest rate pays three dollars for every one hundred dollars kept in the account. The fact that this return (three percent) is measured in terms of dollars is what makes this a nominal interest rate. A return that was measured in terms of goods would be a real interest rate. Because the price level of the economy may change over time, real interest rates and nominal interest rates are not the same, and understanding the difference between the two is important. An example will help illustrate the main issue.

Example:
Consider an economy in which there is only one good – macroeconomics textbooks, say. In the year 2000, the price of a textbook is \$100. Wishing to purchase 5 textbooks, but having no money with which to buy them, you borrow \$500 from a bank. The terms of the loan contract are that you must pay back the principal plus 10% interest in one year – in other words, you must pay back \$550 in one year. After one year has passed, you repay the bank \$550. If there has been zero inflation during the intervening one year, then the purchasing power of that \$550 is 5.5 textbooks, because the price of one textbook is still \$100. Rather than thinking about the loan and repayment in terms of dollars, however, we can think about it in terms real goods (textbooks). In 2000, you borrowed 5 textbooks (what \$500 in 2000 could be used to purchase) and in 2001, you

paid back 5.5 textbooks (what \$550 in 2002 could be used to purchase). Thus, in terms of textbooks, you paid back 10% more than you borrowed.

However, consider the situation if there **had** been inflation during the course of the one year. Say in the year 2001 that the price of a textbook had risen to \$110. In this case, the \$550 repayment can only be used to purchase 5 textbooks, rather than 5.5 textbooks. So we can think about this case as if you had borrowed 5 textbooks and repaid 5 textbooks – that is, you did not pay back any additional textbooks, *even though* you repaid more dollars than you had borrowed.

In the zero-inflation case in the above example, the nominal interest rate is 10% and the real interest rate is 10%. In the 10%-inflation case, however, the nominal interest rate was still 10% but the real interest rate (the extra textbooks you had to pay back) was zero percent. This relationship between the nominal interest rate, the real interest rate, and the inflation rate is captured by the **Fisher equation,**

$$r = i - \pi,$$

where r is the real interest rate, i is the nominal interest rate, and π is the inflation rate. Although almost all interest rates in economic transactions are specified in nominal terms, we will see that it is actually the real interest rate that determines much of macroeconomic activity.

4.4. Distributional Consequences of Unexpected Inflation

Now that we have studied the Fisher equation, we can go back and describe one of the consequences of unexpected inflation that we could not consider earlier. Perhaps the most important consequence of unexpected inflation is the redistribution of wealth that occurs between borrowers and lenders. To illustrate this idea, we will continue our example from above, as well as present an important extension of the idea of the Fisher equation that we just discussed.

The Fisher equation, in its most-often encountered form, actually involves **expected inflation,** which we will denote by π^e, rather than actual inflation. That is, the real interest rate that prevails in the current time period depends on the nominal interest rate that prevails in the current time period and the expected value of inflation between the present and the future, so that we can write the Fisher equation as

$$r = i - \pi^e.$$

Comparing this to the Fisher equation that we presented above, which involved actual inflation π (without superscript), you may be tempted to conclude there is no difference between them. In fact, the difference is an important one. This expectational form of the Fisher equation is the one that economic agents (borrowers and lenders, say) implicitly use when making economic decisions, whereas the non-expectational form is the one that is used to measure real interest rates *after the fact.* We now extend our example from earlier to illustrate this point, as well as describe the distributional consequences of unexpected inflation,

In the example above, you borrowed \$500 from the bank at a nominal interest rate of 10%. Suppose the bank, in determining this nominal interest rate, believed that inflation would be zero between the time you took out the loan and the time you repaid it. That is, the bank's expectation of inflation over the coming year was $\pi^e = 0$. Thus, the bank believed that it would be earning a real interest rate of $r = i - \pi^e = 0.10 - 0 = 0.10$ off its loan to you, which we easily computed using the expectational form of the Fisher equation. In real terms, it expected that it would receive back from you 5.5 textbooks one year after lending you 5 textbooks.

However, suppose after one year has passed, the actual inflation rate turns out to be 10%, so $\pi = 0.10$ (notice the lack of superscript!). We then use the non-expectational form of the Fisher equation to compute the actual real interest rate turned out to be $r = i - \pi = 0.10 - 0.10 = 0$. In real terms, the bank lent you 5 textbooks and then received back only 5 textbooks one year later.

We clearly see that the UNexpected inflation has hurt the lender, the bank, while benefiting the borrower, you. In fact, this is a general conclusion, and a very important one: **inflation which turns out higher than expected hurts lenders and benefits borrowers.**[24] Essentially, the higher-than-expected inflation redistributes real assets (0.5 textbooks in our example) from the lender to the borrower.

[24] Convince yourself, by working through a similar example you construct for yourself, that inflation which turns out lower than expected hurts borrowers and benefits lenders.

5. Aggregate Demand I: Consumption

In discussing the accounting relations surrounding GDP and the ways in which to compute it, we have already become acquainted with the expression

$$GDP = C + I + G + (X - IM),$$

which is perhaps the most straightforward way of computing GDP because good data for each of these components exist for all industrialized countries. The above relationship is an accounting relationship and as such is always true.

We now begin to develop a model of how output and the price level are determined. From introductory microeconomics, you are familiar with the ideas of supply and demand in perfectly competitive markets and the interaction of the two, which determines an equilibrium quantity and price for a particular good. The basic idea in macroeconomic analysis is similar. Rather than considering the demand for or supply of one good at a time, we will consider the **aggregate demand** for and **aggregate supply** of all goods in the economy. The interaction of aggregate demand and aggregate supply determines the equilibrium levels of real GDP and the price level for the economy. You will probably find that you can rely on your intuition from microeconomics regarding demand and supply a good deal – however, the fundamental natures of aggregate demand and aggregate supply differ subtly, but importantly, from their microeconomic counterparts. As such, we will consider the determinants of aggregate demand and aggregate supply in some detail, beginning with aggregate demand.

Consumption is by far the largest component of GDP in the U.S., constituting about two-thirds of GDP. As such, an investigation of aggregate demand should logically begin with a detailed study of the determinants of consumption. Recall from earlier that macroeconomic consumption refers to a hypothetical "aggregate good" which, for the purpose of expressing all of the many individual goods and services that an economy produces in common units, is measured in dollars.

One of the main determinants of consumption is **disposable income,** which is defined as the income available to consumers in some period of time after deducting all taxes and adding all transfers (such as Social Security benefits, veterans' benefits, etc.). Symbolically, $DI = Y - T$, where DI is disposable income, Y is national income (equivalently, GDP), and T represents all taxes and transfers. Historically, real consumption spending has tracked real disposable income very closely in the U.S., as Figure 8 shows.

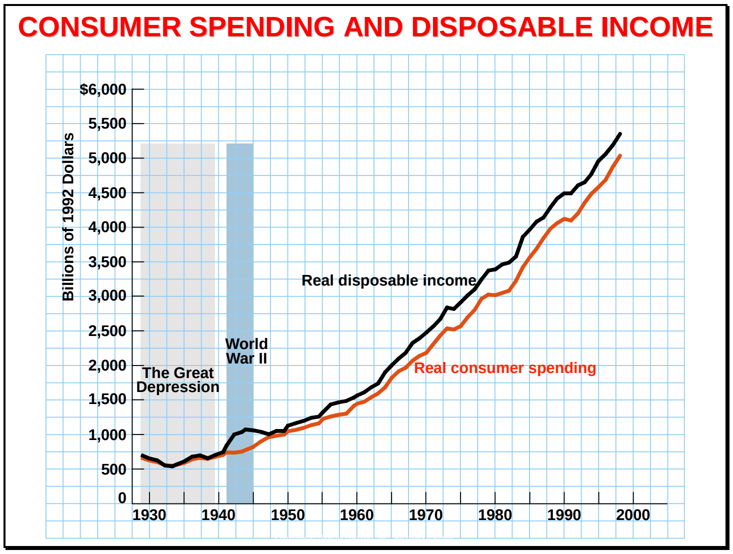

Figure 8

Many alternative models have been developed to relate disposable income, as well as other factors, to consumption. The simplest among these is the Keynesian consumption function.

5.1. Keynesian Consumption Function

The simplest explanation of consumption behavior is that individuals simply spend out of their *current* disposable income, according to a consumption function $C = C(DI)$. For this functional relationship to be able to tell us more about how consumption depends on disposable income, we would need to know the shape of this function. From empirical evidence on real disposable income and real consumption, the relationship seems to be linear – thus, determining the shape of $C(\cdot)$ amounts to finding a slope and a vertical intercept. Again using data for the U.S., the slope of this simple consumption function is approximately 0.9. Figure 9 plots the consumption function for some hypothetical economy.

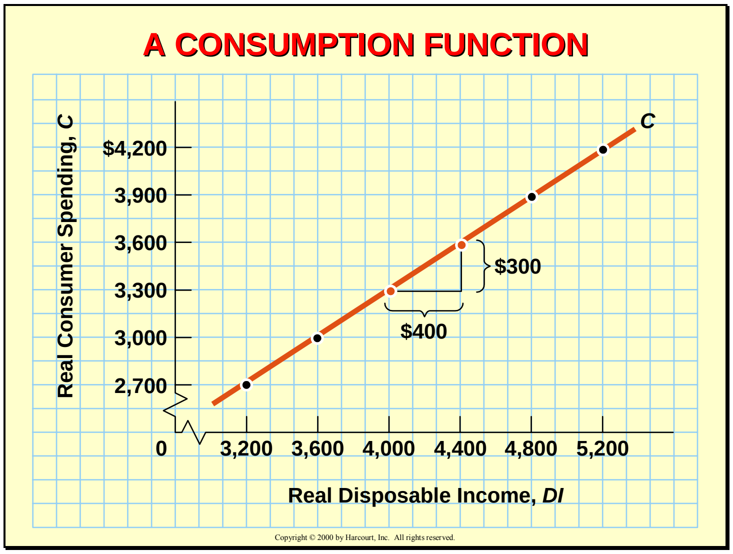

Figure 9

In Figure 9, the slope of the consumption function is 0.75. The slope of the consumption function is known as the **marginal propensity to consume (MPC),** out of disposable income. It simply describes what proportion of every additional dollar of disposable income is spent on consumption. In the economy in Figure 9, 75 cents out of every extra dollar of disposable income is consumed. In the U.S., historical data show that approximately 90 cents of every extra dollar of disposable income is consumed.[25] Note also, although the graph does not clearly show it, that the vertical intercept of the consumption function is not zero. That is, even when disposable income is zero, consumption is strictly positive, presumably because of the ability to borrow, from either financial institutions or family and friends.[26]

5.2. Permanent Income Hypothesis

The Keynesian consumption function, while a useful starting point, is too simple to fully describe consumption behavior. Casual experience should suggest that there are factors besides *current* disposable income that affect consumption. Two important such factors include the amount of wealth that consumers have,[27] and the expectations of future income. Another such factor could be the real interest rate, which theoretically could cause consumers to change their consumption patterns over time. The simple Keynesian

[25] Although some recent estimates put the MPC over the last five to ten years in the U.S. at much closer to one – or even above one.

[26] The study of how such "autonomous consumption" occurs sometimes goes under the rubric "consumption insurance." The terminology is meant to capture the idea that even though one's disposable income is zero, he is able to obtain some positive level of consumption, perhaps by borrowing from friends and family, receiving government benefits, etc. – all of these social mechanisms may loosely be termed "insurance."

[27] Note that wealth and income are two distinct concepts.

consumption function can be modified to include these other determinants, so we will now write the consumption function as

$$C = C(DI; wealth, r),$$

and define wealth as

$$\text{wealth} = \frac{\text{current nominal assets}}{\text{current price level}} + \text{PDV of all future income.}$$

This richer model of consumption determination is often called the permanent income model. Consumption is postulated to depend positively on wealth.[28] However, economic theory actually cannot predict how aggregate consumption will respond to the real interest rate.[29] While the richer consumption function above allows the interest rate as a shift factor, it turns out empirically that the real interest only mildly, if at all, affects consumption. Data show that a rise in the real interest induces a small decrease in current aggregate consumption while a decline in the real interest rate induce a small rise in current aggregate consumption.[30]

In the context of Figure 9, these additional determinants of consumption are shift factors of the consumption function. That is, for example, a change in the price level or a change in the PDV of future income will cause the entire consumption function in Figure 9 to shift.[31]

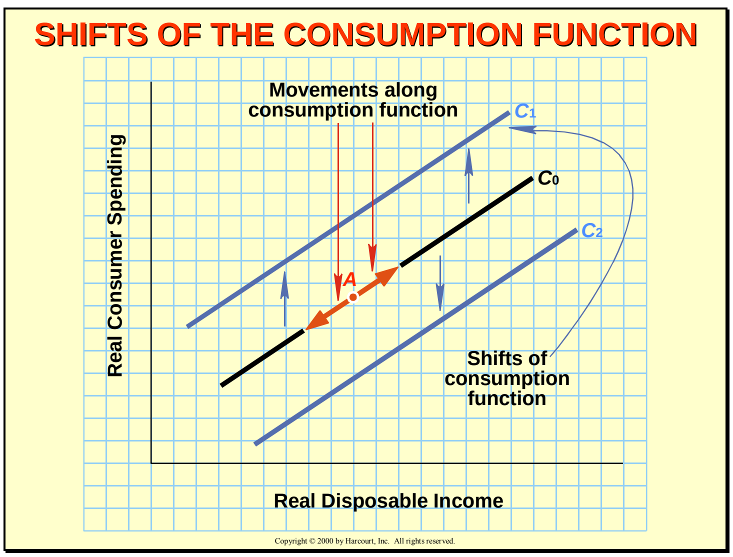

Figure 10

[28] Which should sound natural – if a consumer is wealthier (or even simply *feels* wealthier), she will likely consume more.
[29] The reason why economic theory cannot predict how macroeconomic consumption responds to changes in the real interest rate is left for a more advanced course on macroeconomics.
[30] Economic theory is actually silent on this issue, and that is why we must turn to empirical evidence on this issue. Intermediate microeconomic theory tells us that it is impossible to predict whether overall consumption will rise or fall or be unaffected by changes in the real interest rate.
[31] Recall movements along a curve versus shifts of a curve from microeconomics.

Wealth is an especially interesting determinant of consumption. The "wealth effect" induced by the booming stock market of the mid-to-late 1990's was cited by Federal Reserve Chairman Alan Greenspan as the source of the strong consumption growth seen during that period. Note in particular the way that the first term making up wealth is specified – dividing current *nominal* assets by the current price level shows that current *real* assets affect wealth.[32] Thus, current real assets depend on the price level – and hence consumption depends on the price level. For example, an increase in the current price level reduces current real assets (holding current nominal assets fixed), thus reducing wealth, which in turn reduces consumption.

In Figure 10, then, an increase in the price level would shift the consumption function downwards – that is, for any possible level of disposable income, consumption would fall if the price level rose. We see this relationship between the price level and consumption in **Error! Not a valid bookmark self-reference.** and call this relationship **aggregate consumption demand.**

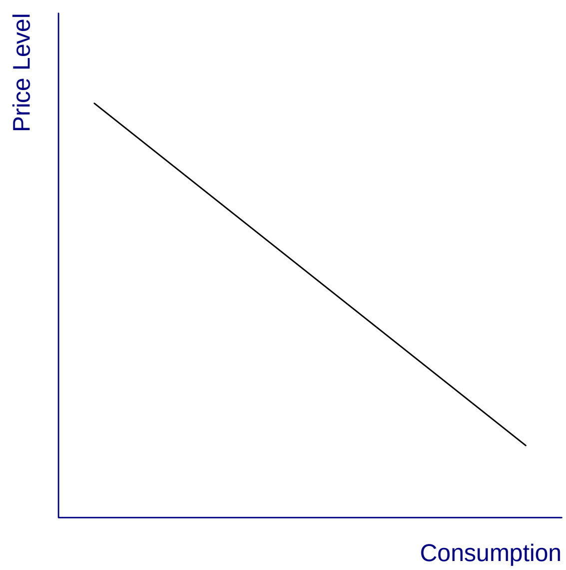

Figure 11

[32] Recall the distinction between nominal and real variables made already in the contexts of GDP and interest rates. The idea is the same here.

6. Aggregate Demand II: Investment

In the previous section, we studied macroeconomic consumption. In this section, we turn to the study of another of the major components of GDP, investment. Right away, it is crucial to distinguish the macroeconomic notion of investment from one that may be more familiar from colloquial usage. Macroeconomic **investment** is the sum of business purchases of **capital goods** (goods that businesses use in the production of other goods and services), new homes built, and addition to firms' inventories. In everyday language, we often use the term "investment" to refer to someone's collection of stocks and bonds, as in "I invested in 100 shares of Microsoft stock last week." In formal economics, this latter type of activity is **not** investment. In fact, this latter type of activity is termed savings, a topic we will encounter later. It is very important, however, to keep this terminology straight, as it is often a source of confusion when discussing matters of savings and investment.[33]

Because it is consumers (for the most part) that purchase homes, we see from the above definition of investment that investment encompasses activities of both consumers and firms. However, for convenience of exposition, we will simply speak of investment as being undertaken by firms only.

In the following, we will consider the benefits and costs facing a single firm when it is faced with investment opportunities. We will then appeal to a basic result from microeconomics, that optimal decisions imply that marginal benefit equals marginal cost, to determine the optimal amount of investment this single firm should undertake. From this, we will extrapolate to firms in aggregate so that we can speak of investment demand in the entire economy.

6.1. Benefits of Investment

Recall our notion of the production function, $Y = AF(K, N)$, where K stands for capital. Capital goods are goods that businesses use in the production of other goods and services, for example, the trucks used by a delivery company, the hair dryers used in a beauty salon, and the manufacturing plants used to assemble cars. Such goods naturally wear out over time. The wearing out of capital goods is termed **depreciation.**[34]

In the U.S. data show that roughly eight percent of the nation's capital stock depreciates every year and that this number has been fairly stable for a long period. Studies from

[33] As we will learn later, there is in fact a deep connection between macroeconomic savings and macroeconomic investment. For now, however, we are only considering the topic of investment.

[34] This idea of wearing out of goods is the economic notion of depreciation, and has nothing to do with any types of depreciation rules you may have learned in an accounting class. Accounting standards and regulations are such that sometimes a company has some control over how to *report* its depreciation of capital goods (i.e., accelerated depreciation, straight-line depreciation). Our economic notion of depreciation has only to do with how quickly goods *actually* wear out over time and nothing to do with how a company may choose to *report* how quickly goods wear out.

other countries show that their depreciation rates are also fairly stable over time, though they may differ from the U.S. depreciation rate. We thus adopt the assumption that there is constant rate of depreciation each year and we denote this constant yearly depreciation rate by δ (the Greek letter "delta"), where $0 < \delta < 1$. Thus, if a business own K units of capital at the beginning of the year and purchases no capital goods at all during the year, at the end of the year it will own $(1-\delta)K$ units of capital.

We now introduce two more pieces of terminology, gross investment and net investment. Suppose the business just mentioned was a "hair-drying salon" (all they do is dry hair!), and the only capital goods it uses is hair dryers. Say the business did purchase some new hair dryers during the course of the year. Specifically, suppose it purchased x units of hair dryers. We would call this absolute quantity of capital goods purchased **gross investment.** But it already had some hair dryers to begin with, some of which wore out during the course of the year. At the end of the year, the business would have $(1-\delta)K + x$ hair dryers. We see that even though the salon purchased x hair dryers during the year, the net addition to the number of hair dryers during the course of the year was actually less than x due to depreciation. **Net investment** is gross investment minus total depreciation. In our example, total depreciation is δK, which represents the number of hair dryers that wore out. So we have in our example that gross investment by the salon was x while net investment was $x - \delta K$.

Now let's turn to the benefits of investment. The reason why our salon purchases hair dryers is because it helps them conduct business – it helps them to produce output, in other words. Recall from our initial look at the aggregate production function that total output displays diminishing marginal product in capital (as well as in labor), which is captured graphically by the fact that the slope of the curve in Figure 12 is always declining.

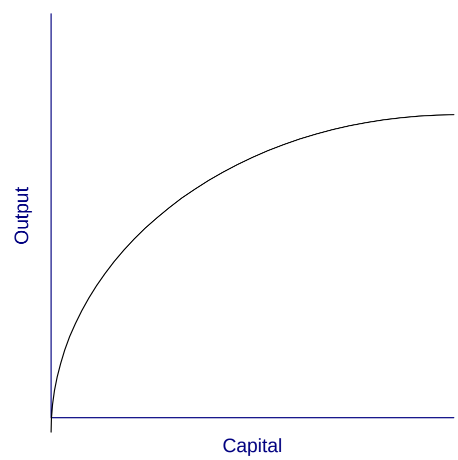

Figure 12

Appealing to your basic knowledge from microeconomics, the reason that a firm (our salon) would want to purchase *extra* capital is because of the *extra* output that it could produce with it. In more formal terminology, it is the **marginal product of capital** that is the benefit to a firm of purchasing capital – how much extra output a firm can produce by purchasing extra capital. This marginal product is diminishing as capital increases, as seen in Figure 13 ,which is simply a graph of the slope of the curve in Figure 12.

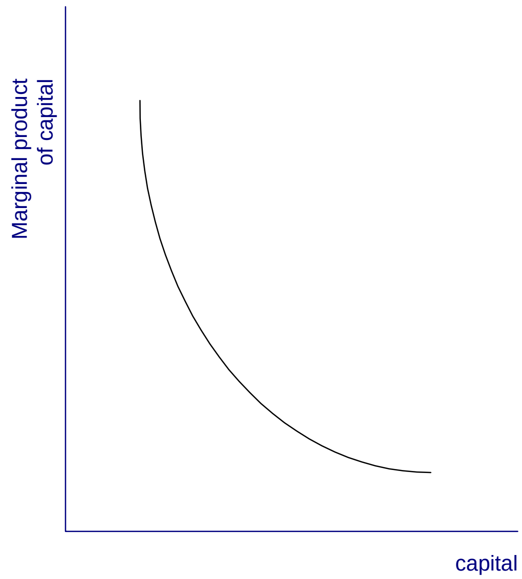

Figure 13

6.2. Costs of Investment

The relevant cost of investment is a little more subtle to recognize. The relevant cost of investment is *not* the price of the capital goods being purchased. Rather, it is the *opportunity cost* to the firm. Continuing our hair-drying salon example, rather than purchase new hair dryers, the firm could alternatively have put its funds in some sort of savings device (a bank account that pays interest, say) and earned interest on it. The relevant interest rate is the real interest rate, rather than the nominal interest rate that the account specifies. But we can easily compute the real interest rate from the nominal interest rate using the Fisher equation if we know something about the inflation rate.[35] By buying hair dryers rather than putting its money in some interest-bearing account, the salon is forgoing the interest it could have earned – hence the real interest rate is the relevant cost of investment.

The real interest rate is independent of the amount of capital a firm owns, in constrast to the way in which the marginal product of capital depends on the amount of capital (see Figure 13). Thus, we have Figure 14.

[35] If you're unsure about the mechanics or intuition behind the Fisher equation, now is the time to review that topic.

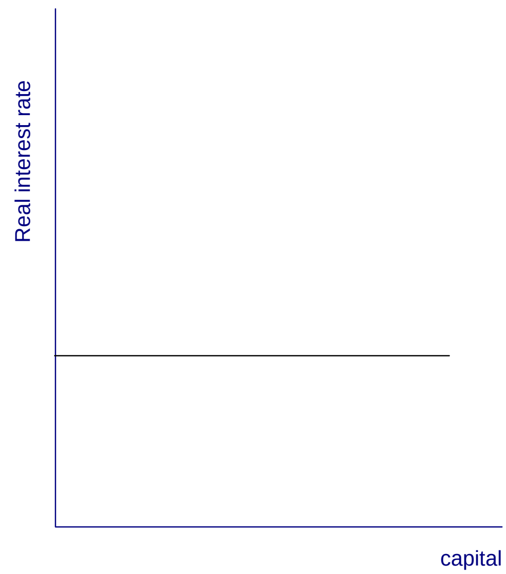

Figure 14

6.3. Optimal Investment

To determine optimal investment by a firm, we put Figure 13, which describes the benefit of investment, together with Figure 14, which describes the cost of investment, to generate Figure 15.

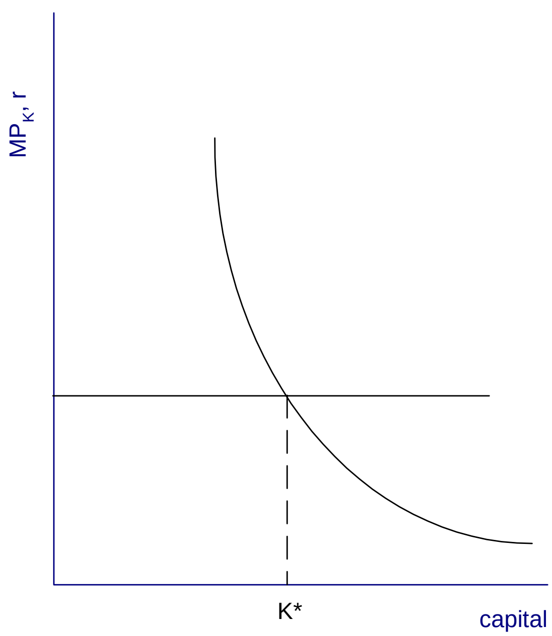

Figure 15

In Figure 15, the quantity K^* denotes the optimal quantity of capital the firm would like to have. This quantity of capital is optimal because it is the amount of which the marginal benefit of capital (the marginal product) equals the marginal cost of capital (the real interest rate). A basic lesson from microeconomics is that optimal decisions occur at points at which marginal benefits balance marginal costs – this is exactly the principle here.

We are not quite finished, however. Figure 15 describes the optimal amount of capital – we are interested in determining the optimal amount of investment. With the terminology we developed at the beginning of this section, it is easy to make this connection.

Suppose the firm currently has K units of capital, and its target level of capital one year from now is K^*, determined through a diagram like Figure 15. Further suppose that $K < K^*$. Then Figure 16 shows us the amount of gross investment as well as net investment the firm must undertake.[36]

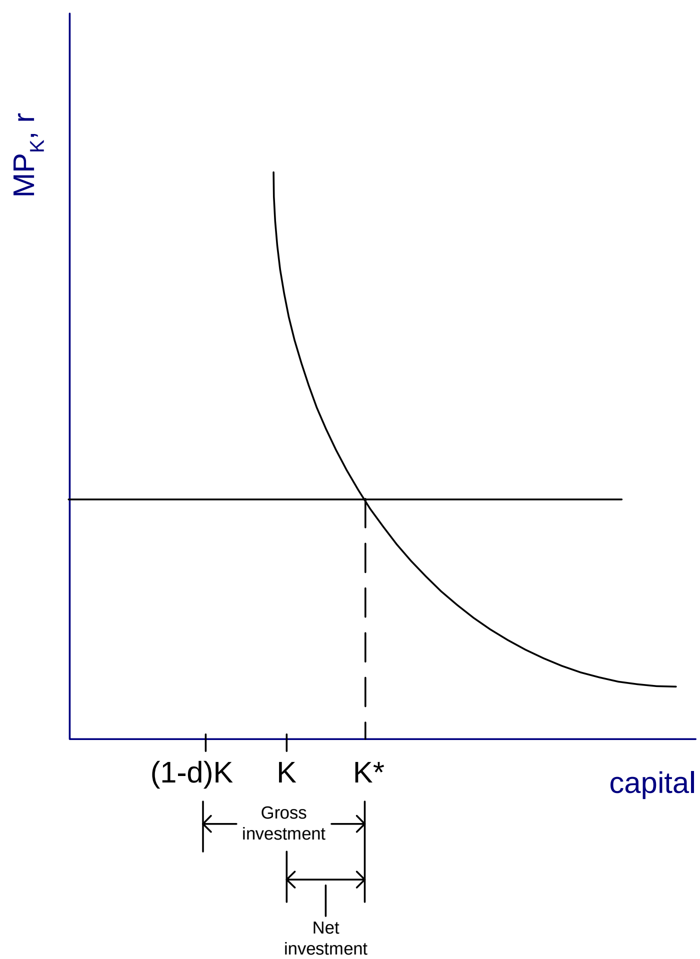

Figure 16

[36] A useful exercise is to confirm for yourself the relationships shown in Figure 16 using the definitions of investment we introduced earlier.

6.4. Macroeconomic Investment Demand

Now suppose there is a large number of firms just like the one we have studied, each of which faces the same costs and benefits of investment. Each firm would make optimal investment decisions according to the same principles. As we have described, the relevant cost of investment is the real interest rate. Now we will trace out the effect on investment (whether net or gross) of changed in the real investment. Doing so will generate the **investment demand** function.

Our analysis here uses Figure 16. Suppose that the real interest rate decreased, while current capital K, the depreciation rate δ, and the marginal product of capital function all remained unchanged. As Figure 17 shows, this would lead to a larger desired level of optimal capital, named K^{**} in Figure 17, and hence larger amount of optimal (both gross and net) investment, as can be seen by comparing with Figure 16.

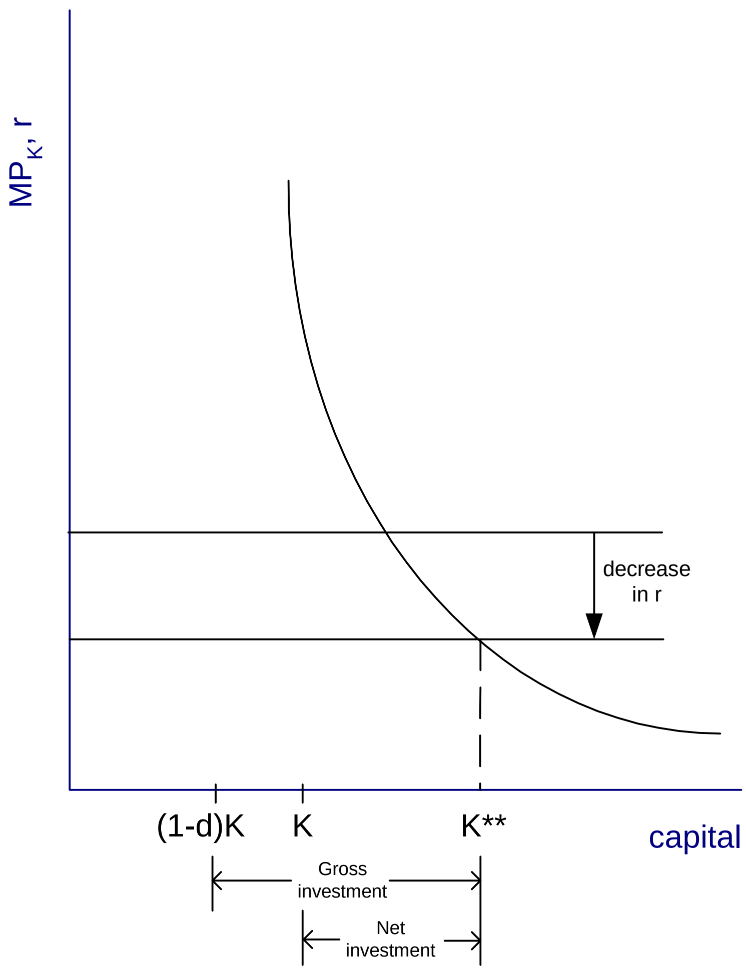

Figure 17

Similarly, a rise in the real interest rate would lead to a decline in investment. This relationship is captured in Figure 18, which is the macroeconomic investment demand curve. Notice that here we have finally put investment, rather than capital, on the axes of our graph.

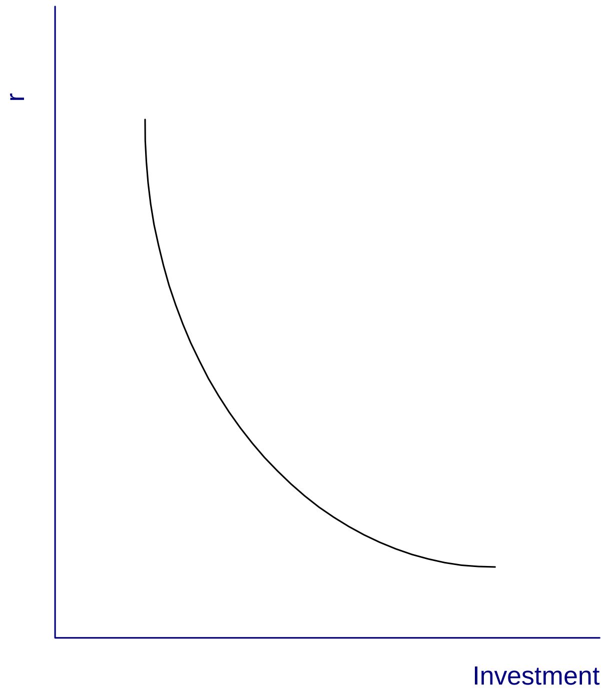

Figure 18

7. Aggregate Demand III: Demand-Side Equilibrium

We have considered the major determinants of consumption (current disposable income and wealth) and investment (the marginal product of capital and the expected real interest rate), learning that real GDP (equivalently, real national income) affects consumption decisions but not investment decisions. That is, national income affects aggregate consumption because disposable income affects consumption – but disposable income is simply national income less taxes. Ignoring the wealth channel for the moment, we can thus write consumption as a function of disposable income,

$$C = a + b(Y - T),$$

where $(Y - T)$ is disposable income, a is the vertical intercept of the consumption function, and b is the slope of the consumption function. Recall that the slope b is the **marginal propensity to consume** out of disposable income. We adopt this linear form for the consumption function simply because, as already discussed, empirical observation of the U.S. economy suggests it. Using this specification, wealth is clearly a shift factor of the consumption function, and hence, in particular, the price level is a shift factor of the consumption function.

Empirical data shows that government spending and net exports respond very little to real GDP. There are of course many factors that influence these components of spending, but (current) real GDP is not one of them. In this section, we combine the components of GDP to develop the theoretical notion of aggregate demand.

7.1. Aggregate Expenditure

We define **aggregate planned expenditure** for an economy as:

$$AE = C + I + G + NX,$$

where AE stands for aggregate expenditure**.** Simply put, aggregate expenditure is the sum of all (net) expenditures on domestically-produced goods in a given period of time.

The above definition looks very much like the national income identity, $GDP = C + I + G + NX$ which was encountered before. Hence, it seems natural to ask whether this implies that real GDP (equivalently, national income) and aggregate expenditure are simply the same concept. The answer here is subtle. It is true that in practice, the way GDP is often computed for an economy is to simply sum up the components according to the national income identity. However, the implicit assumption underlying this procedure is that real GDP *equals* aggregate expenditure in some given period of time. This need not be the case, however. It is entirely possible that aggregate expenditures in an economy during, say, a given year, are less than the real GDP for that year. The difference ends up as excess inventories for firms.

To illustrate the above, think of a simple economy in which there is no government spending, no investment, and no exports or imports. Furthermore, there is only one good in this economy – cars, say. If in a given year, car manufacturers produce 10,000 cars, but customers only purchase 9,000 cars, the excess 1,000 cars wind up sitting on dealers' lots unsold, ready to be sold to customers in the future. In this particular year, then, inventories have increased. This increase in inventories provides a strong signal to manufacturers, however, that demand for cars is not as strong as they had thought. Thus, in the next year, it is likely they will cut back on production. As long as inventories continue to rise, this signal will remain intact and producers will cut back on production. Eventually, this adjustment process will have proceeded to the point where the production of cars is equal to aggregate planned expenditures on cars. This latter situation is **demand-side equilibrium,** which occurs only when aggregate expenditure equals real GDP.[37] Thus, the use of the national income identity to compute GDP implicitly assumes that the economy is in demand-side equilibrium.

7.2. Income-Expenditure Diagram

Using the definition of aggregate expenditure and the simple consumption function above, we can derive the **aggregate expenditure function** through some simple manipulation:

$$\begin{aligned} AE &= C + I + G + NX \\ &= C(DI) + I + G + NX \\ &= a + b(Y - T) + I + G + NX \\ &= a + bY - bT + I + G + NX. \end{aligned}$$

Note that the term multiplying GDP in this last expression, which is simply the marginal propensity to consume, is smaller than one. Thus, aggregate expenditure is a linear function of real GDP, with a slope strictly smaller than one. Figure 19 plots the aggregate expenditure function.

[37] You should convince yourself that the mechanism works analogously in the other direction, as well – that is, from a point at which aggregate expenditure is larger than real GDP.

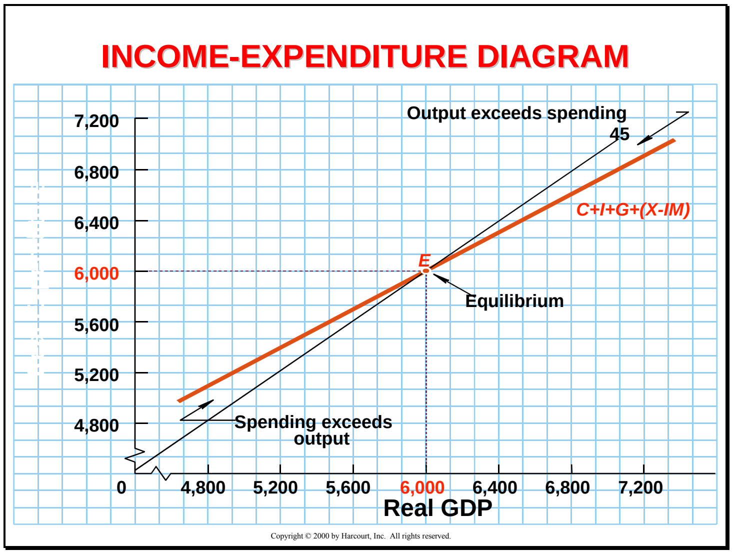

Figure 19

The heavy line in Figure 19 is the aggregate expenditure function, and the lighter line is real GDP plotted as a function of real GDP, which of course simply results in a 45-degree line. This 45-degree line is included to highlight where demand-side equilibrium occurs. Graphically, demand-side equilibrium occurs where the 45-degree line crosses the aggregate expenditure function, which accords with the illustration of the concept in the example above. Because the 45-degree line has a vertical intercept of zero and a slope of one, and because the aggregate expenditure function has a strictly positive vertical intercept and a slope smaller than one, there must indeed be a unique intersection of the two functions.

7.3. The Aggregate Demand Curve

In the simple description of the consumption function above, real GDP (disposable income) is posited to be the only determinant of consumption. Recall, however, the richer specification of the consumption function described earlier – specifically, that consumption depends on wealth as well. Looking one level deeper into this richer consumption function, we saw that the price level affected consumption through the wealth channel.

Consider the relationship just discussed between aggregate expenditure and real GDP, specifically $AE = C(DI) + I + G + NX$. Because the current price level does not affect investment, government spending, or net exports,[38] the only channel through which the current price level affects aggregate expenditure is through consumption. We have already seen that an increase in the price level shifts the entire consumption function downwards, and a decrease in the price level shifts the entire consumption function

[38] To a first approximation, at least.

upwards. Thus, the effect of the price level on the aggregate expenditure function will be qualitatively the same: an increase in the price level shifts the aggregate expenditure function downwards, and a decrease in the price level shifts the aggregate expenditure function upwards. This effect is illustrated in Figure 20.

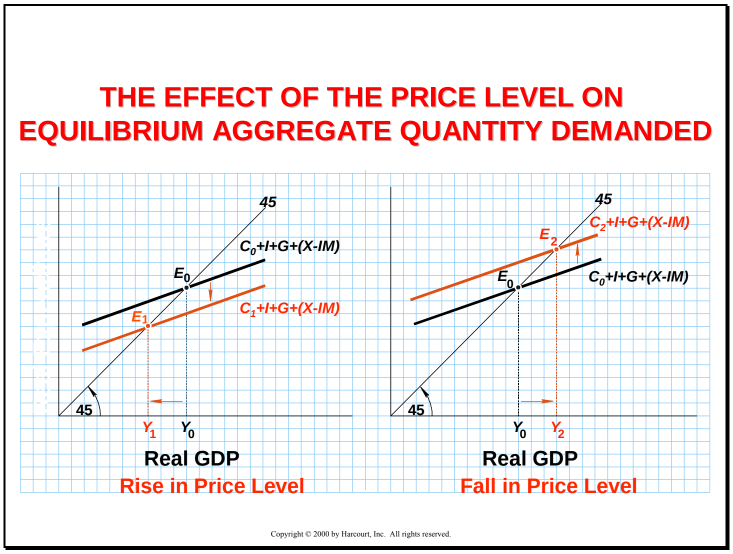

Figure 20

Note especially that the point of demand-side equilibrium GDP moves in the opposite direction as the price level. That is, as the price level increases, the demand-side equilibrium level of real GDP falls, and as the price level decreases, the demand-side equilibrium level of real GDP rises. This relationship is captured by the **aggregate demand curve**, which is a graph of the price level on the vertical axis and real GDP on the horizontal axis, as Figure 21 shows.

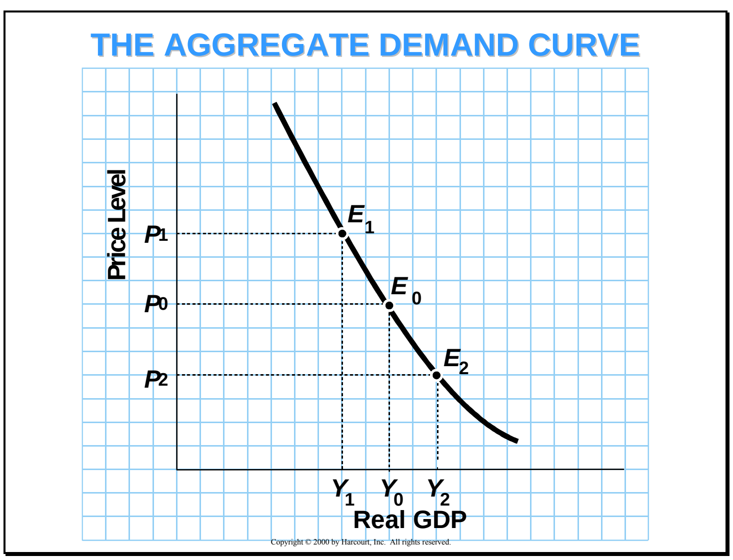

Figure 21

The aggregate demand curve surely looks similar to the usual downward-sloping demand curves you are familiar with from microeconomics. However, note the important conceptual difference between the aggregate demand curve and a demand curve representing demand for one particular good. The inverse relationship arises at the macroeconomic level because of aggregate consumption's dependence on wealth, and wealth's dependence on the price level. The inverse relationship at the microeconomic level arises simply because of diminishing marginal utility. It is **not** correct to interpret the inverse relationship that arises at the macroeconomic level as a situation in which, in aggregate, a higher overall price level induces lower expenditures simply because goods and services have become more expensive.[39]

Thus, the aggregate demand curve is a relationship between the price level and aggregate quantity demanded when the economy is in demand-side equilibrium.

7.4. The Keynesian Multiplier

In demand-side equilibrium, $Y = AE$. Using the aggregate expenditure function developed above, the demand-side equilibrium relationship can be expressed as

$$\begin{aligned} Y &= AE \\ &= a + bY - bT + I + G + NX. \end{aligned}$$

Notice that real GDP, Y, now appears on both sides of the equality. Solving for Y therefore yields

$$Y = \frac{1}{1-b}\left[a - bT + I + G + NX\right].$$

Expanding this expression once again yields

$$Y = \frac{a}{1-b} - \frac{b}{1-b}T + \frac{1}{1-b}I + \frac{1}{1-b}G + \frac{1}{1-b}NX,$$

which reveals an important relationship between real GDP, taxes, investment, government spending, and net exports. The terms multiplying taxes, investment, government spending, and net exports on the right-hand side of the previous expression are each component's **multipliers.** These multipliers describe how a change in one of these components affects demand-side equilibrium real GDP. For example, a one unit increase in investment leads to an increase in real GDP of one unit *times* $1/(1-b)$. Because the marginal propensity to consume, b, is between zero and one, $1/(1-b)$ is obviously larger than one. Many empirical studies using U.S. data for the past several

[39] Indeed, this important conceptual distinction was already made when the consumption function was initially discussed. With our assumptions that investment, government spending and net exports are unaffected by real GDP, the aggregate expenditure function inherits the (conceptual) properties of the consumption function.

decades have reached the conclusion that the marginal propensity to consume is approximately 0.9. This implies that the multiplier associated with investment is $1/(1-0.9)=10$.

As another example, consider the multiplier on taxes. Because this multiplier is negative, a one unit increase in taxes leads to a *decrease* in demand-side real GDP of $b/(1-b)$, which, using the empirical value $b=0.9$, yields a multiplier on taxes in the U.S. of –9 (Note the negative sign!). In this simple formulation, government spending and net exports are seen to have the same numerical multipliers as the one on investment.[40]

The presence of the multiplier can be seen graphically in Figure 19. There, demand-side equilibrium occurs at a level of $Y=6,000$. Consider an increase in government spending of, say, one unit. Such an increase in G shifts the entire aggregate expenditure function upwards by one unit, leading to a larger demand-side equilibrium level of real GDP. For the economy to reach this new level of GDP, however, the economy must travel along the new higher aggregate expenditure curve (given that the economy starts at $Y=6,000$). Because the slope of the aggregate expenditure curve is smaller than one, (in particular, the slope is b), arriving at the new demand-side equilibrium GDP necessarily means, in this case, traveling $1/(1-b)$ units to the right on the horizontal axis. That is, demand-side equilibrium GDP rises by $1/(1-b)$ units, not just the one unit which was the assumed initial increase in government spending. Using empirical values for the U.S., the new demand-side equilibrium real GDP in Figure 19 would be $Y=6,010$. In Figure 21, therefore, the one unit rise in government spending would lead to a rightward shift of the aggregate demand curve by ten units.

The basic economic mechanism at work is simple, yet important. one individual's (or institution's) *expenditure* is another individual's *disposable income*. This second individual, now with more disposable income, will be induced to spend some fraction (the MPC) of it – which winds up as some third individual's disposable income. This third individual, now with more disposable income, will be induced to spend some fraction (the MPC) of it – which winds up as some fourth individual's disposable income. This mechanism proceeds unendingly – the end result of these multiple rounds is that demand-side equilibrium real GDP increases by more than the initial increase in aggregate expenditure. Thus, through these multiple rounds of consumption, the effect of the initial increase in spending is magnified through the economy.[41]

[40] Richer models allow for government spending and net exports to depend on real GDP as well. In those models, there would be additional terms (intercepts and slopes) that would appear in the multipliers, rendering the different multipliers unequal. Recall that as a first approximation, though, government spending and net exports do not vary much with real GDP.

[41] Technical note: This simple multiplier is based on an "infinite" number of rounds of expenditures, and is based on the formula for an infinite geometric series, which you may recall from mathematics:

Empirically, the multiplier associated with government spending in the U.S. is estimated at approximately 2 – much lower than the simple $1/(1-b)$ expression would have us believe. One reason for the discrepancy is the simple "infinite rounds of new expenditure" mechanism used in computing the theoretical multiplier (see the technical footnote above). In reality, of course, "infinite rounds" of spending cannot exist. A more important reason for the discrepancy is that the analysis conducted so far does not consider aggregate supply at all. While it is true that the aggregate demand curve will shift out by $1/(1-b)$ times the change in government spending (assuming again the "infinite rounds of expenditure" simplification), equilibrium GDP (yet to be defined, and not to be confused with demand-side equilibrium GDP) will rise by less as long as the aggregate supply curve is upward-sloping. Thus, a fuller analysis of the empirics of the multiplier must be postponed until aggregate supply has been discussed. Nonetheless, the concept of multipliers affecting aggregate demand is a very important one from a policy perspective – indeed, it is what allows policy-makers to speak of (whether they understand the details or not) macroeconomic policy being able to "prime the pump," so to speak, of the economy.[42]

7.5. Recessionary and Inflationary Gaps

Two important concepts that will arise again once we have considered the interaction of aggregate demand with aggregate supply are **recessionary gaps** and **inflationary gaps.** *For the purpose of this brief introduction to these concepts, assume that demand-side equilibrium GDP **is** the actual equilibrium GDP.* With this assumption, a recessionary gap, simply stated, occurs when actual equilibrium GDP is lower than potential GDP[43] and an inflationary gap[44] occurs when actual equilibrium GDP is higher than potential GDP.

When we introduce aggregate supply into the model and allow it to interact with aggregate demand, it will become much clearer what these concepts actually are – for the moment, consider this just an introduction to terminology.

7.6. So How Valid is the National Income Identity?

$$\sum_{i=0}^{\infty} xy^i = x + xy + xy^2 + xy^3 + \dots$$
$$= \frac{x}{1-y}$$

[42] You should convince yourself that the arguments all work analogously for decreases in any of the components of aggregate expenditure.

[43] Recall that potential GDP is defined as that level of GDP associated (through the aggregate production function) with full employment.

[44] The explanation for the terminology "inflationary" will have to be deferred until we have properly considered aggregate supply.

One of the first lessons of the course was that GDP is computed according to the national income identity, $Y = C + I + G + NX$. Here, we have seen that, in fact, GDP need not equal the sum of consumption, investment, government spending, and net exports. That is, real GDP in any given time period need not equal aggregate expenditure. The difference, if there is one, is reflected in a change in inventories held in the economy. While a full discussion of inventory management is beyond the scope of this text, the maintained assumption we will use throughout our analysis is that firms desire to keep their stock of inventories constant. This maintained assumption implies that the economy is always in demand-side equilibrium.

Data on inventory fluctuations for the U.S. show that they are very small fractions of GDP. Thus, our maintained assumption is a good approximation and indeed validates the approach to computing GDP as simply the sum of the components of aggregate expenditure.

8. Aggregate Supply

Thus far, we have considered aggregate demand in isolation. In addition, we have briefly considered the impact of fiscal policy on the demand side of the economy – in particular, a change in government spending or lump-sum taxes shifts the entire aggregate expenditure schedule and thus the entire aggregate demand curve. The aggregate demand curve shifts by an amount equal to the amount of the change in government spending or taxes *times* the appropriate multiplier. The magnitude of the shift in the aggregate demand curve is precisely equal to the change in the demand-side equilibrium real GDP induced by the particular fiscal policy and described by the income-expenditure diagram.

However, the effects of such policy actions cannot be properly traced through the macroeconomy unless we also consider the supply side of the economy, a task to which we now turn. In the discussion that follows, a very important asymmetry becomes evident between the theory of aggregate supply and the theory of aggregate demand. Specifically, we are able to distinguish two different notions of aggregate supply – a short-run notion of aggregate supply and a long-run notion of aggregate supply – whereas there is only one notion of aggregate demand.

8.1. Short-Run Aggregate Supply

As you are familiar with from basic microeconomics, the objective of firms is to maximize profits. For simplicity, assume that labor is the only input used by firms to produce output. This assumption is not a bad one for a short-run analysis because, as we have already seen, capital is relatively fixed in the short-run. Thus, the objective of firms is to maximize

$$\begin{aligned} \text{Profit} &= \text{Revenue} - \text{Cost} \\ &= P \cdot Q - W \cdot N \end{aligned}$$

where P is the price level of output (i.e., P is the aggregate price level), Q is the quantity of output goods produced, W is the nominal wage rate,[45] and N is the quantity of labor hired. In this one-input environment, the production function can be expressed as $Q = f(N)$, and $f(N)$ displays diminishing marginal returns.[46] So the firm's objective is

$$\max_N \left\{ P \cdot f(N) - W \cdot N \right\}.$$

[45] That is, it is the wage denominated in terms of currency (as opposed to the *real* wage rate, which is denominated in terms of actual goods).

[46] This is no different from using the Cobb-Douglas production function we have been considering with the capital stock held fixed at constant level. This approach basically just saves on notation.

The term $P \cdot f(N)$ is simply total revenues, and the term $W \cdot N$ is simply total costs.[47] Graphically, the objective of the firm is to choose that level of labor input which maximizes the vertical distance between the total revenue curve and the total cost curve:

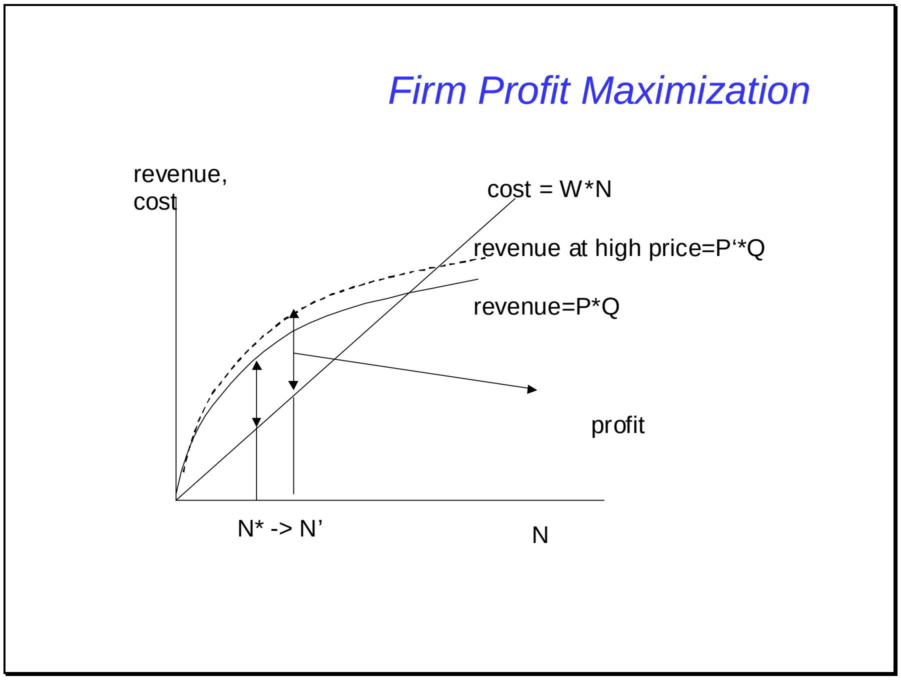

Figure 22

Figure 22 shows two total revenue curves, one for a low price of output goods and one for a high price of output goods,[48] and it also shows the cost curve under the assumption that labor is the only input to production. **Note that the cost curve here is linear. The linear cost curve arises because of the assumption that the nominal wage rate is fixed in the short-run.** The nominal wage rate is often considered to be fixed in the short-run because most employees have some kind of labor contract (explicit or implicit) that fixes their nominal wage for some period of time. Thus, even if the general price level rises, workers' nominal wages are fixed for some time simply due to their labor contracts.[49]

At the new higher price level in Figure 22, which *excludes* the nominal wage rate because P is the price level of output, firms will find it optimal to hire more labor input than before, which, by the production function $Q = f(N)$, implies that output rises. Thus, a higher price level leads to increased production by firms – in other words, a higher price level is associated with increased aggregate supply. This relationship is represented by the upward-sloping aggregate supply curve in Figure 23.

[47] With the full Cobb-Douglas specification of the production function, of course, total costs would also include the rental cost of capital, not just the cost of labor.
[48] The price of output goods should simply be thought of here as the price level *excluding* wages.
[49] This of course means that workers' *real* wages decrease if the price level rises.

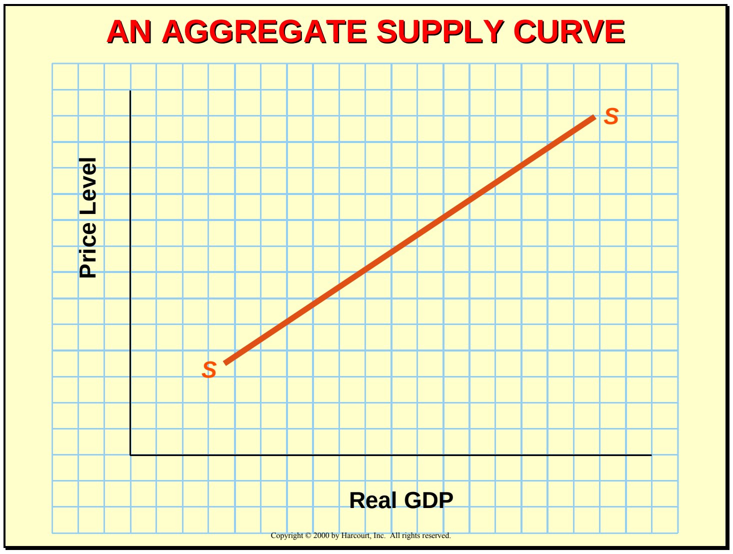

Figure 23. The short-run aggregate supply curve shows a positive relationship between the price level and real GDP.

It is important to emphasize again the assumption used to derive the upward-sloping short-run aggregate supply curve: the nominal wage is assumed to not change with the rest of the price level of the economy. This idea is often called the **sticky wage** hypothesis, which captures the (somewhat empirically-supported) idea that wages are slower to adjust than prices of goods and services because of the presence of labor contracts.[50]

8.1.1. Shift Factors of the Short-Run Aggregate Supply Curve

The short-run aggregate supply curve is a relationship between quantity supplied by firms and the price level exclusive of wages *holding all other determinants of quantity supplied constant* (just as the aggregate demand curve is a relationship between quantity demanded by households and the price level *holding all other determinants of quantity demanded constant*). Thus, a change in any of these other determinants of quantity supplied will cause a shift of the aggregate supply curve. There are four main shift factors of the short-run aggregate supply curve, which are:

1. The money wage rate (can rederive the firms' profit-maximizing level of labor input N to see that output will change *for a given output price P*).
2. Materials prices (i.e., an increase in the price of other inputs besides labor)
3. Total factor productivity (the level of technology, which we denoted by A in our study of growth accounting).
4. Shifts in the supplies of labor and capital (more generally, in the supplies of any inputs to production)

[50] As well as other labor market institutions, such as unions and simple psychological resistance to downward wage revisions.

The short-run aggregate supply curve shifts *inwards (leftwards)* due to an increase in the money wage rate, an increase in materials prices, a decrease in total factor productivity, and decreases in the supplies of labor and capital. The short-run aggregate supply curve shifts *outwards (rightwards)* due to a decrease in the money wage rate, a decrease in materials prices, an increase in total factor productivity, and increases in the supplies of labor and capital.

Consider in particular an increase the nominal wage rate. The effect on the aggregate supply curve is illustrated in Figure 24. The reason for the leftward shift is simply that for any given price level, producers will produce less due to the increase in their per-unit labor costs.

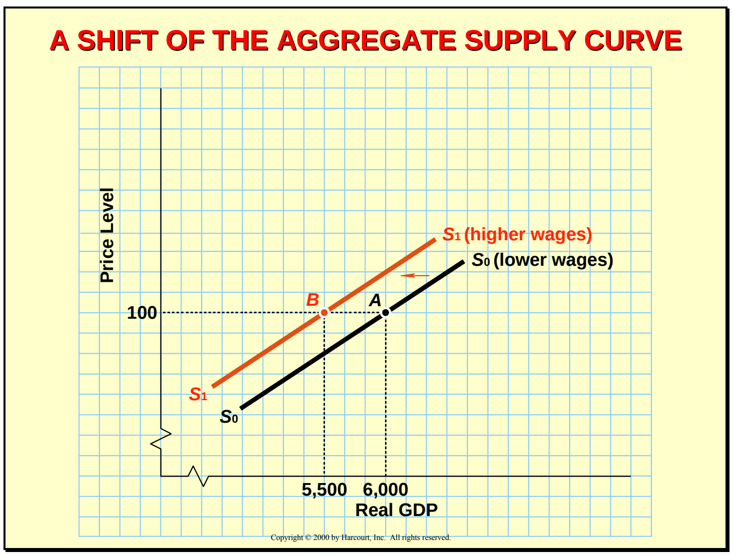

Figure 24

8.2. Long-Run Aggregate Supply

Recall that in the derivation of the aggregate demand curve, we specifically *included* the nominal wage rate as part of the aggregate price level. We then determined that the negative relationship between the aggregate price level and aggregate demand arose because of the wealth channel in the consumption function. However, in the derivation of the short-run aggregate supply curve immediately above, the nominal wage rate was intentionally omitted from the price level by appealing to the sticky wage theory.

Labor contracts are not fixed forever, of course. They periodically are subject to revision. If the price level has risen between these periodic revisions, workers will naturally seek increases in their wages when their contracts are up for renegotiation. Assuming some form of marginal product pricing in the labor markets, firms will have to accommodate workers' requests for higher nominal wages – which leads to a leftward shift of the short-run aggregate supply curve exactly as in Figure 24. The increase in nominal wages induces firms to cut back on the amount of labor they hire (in terms of Figure 22, the slope of the cost line increases), which leads to a reduction in aggregate supply. Because the nominal wage is not on the axes of the graph in Figure 24, the entire schedule shifts to the left.

In equilibrium, the renegotiated nominal wages rise by exactly the same percentage as the price level – that is, nominal wages are (eventually) adjusted according to the rate of inflation in the economy. Thus, again referring to Figure 22, the amount of labor that firms will hire will return to the *original* profit-maximizing amount of labor, denoted there by N^*. This then implies that the aggregate quantity produced by firms is exactly what it was before the rise in output prices. **Thus, in the long-run, the aggregate supply curve is vertical. In particular, the long-run aggregate supply curve is vertical at the level consistent with full-employment of the economy.** In other words, the long-run aggregate supply curve is vertical at the level of potential GDP.

One final important point is in order about the long-run aggregate supply curve. In the static analysis being presented here, potential GDP is considered to be constant. In an actual economy, however, potential GDP changes over time as well, primarily due to changes in total factor productivity (TFP). Thus, increases in TFP shift the long-run aggregate supply curve rightwards, while decreases in TFP shift the long-run aggregate supply curve leftwards.

8.3. Reconciliation of Long-Run and Short-Run Aggregate Supply?

At this point, you should be somewhat troubled by the asymmetric treatment of wages in the derivation of aggregate demand and in the derivation of aggregate supply. In the derivation of aggregate demand, the nominal wage rate was always described as being included in the price level of the economy. Indeed, that is why exclusive emphasis was placed on the wealth channel as the reason for the negative relationship between the price level and aggregate demand. This mechanism is in keeping with the permanent income hypothesis.

However, in the derivation of aggregate supply, the nominal wage rate is assumed to be excluded from the price level in the short-run (the sticky wage hypothesis) in order to generate an upward-sloping aggregate supply curve. In the long-run, nominal wages are assumed to fully adjust to the price level, generating the vertical long-run aggregate supply curve.

This asymmetric treatment of nominal wages in the theory of aggregate demand and the theory of aggregate supply is in fact an unsettled area of macroeconomic theory. If the nominal wage was always included in the price level, then the aggregate supply curve should always be vertical at the level of potential GDP. Indeed, this would render the theory of aggregate supply symmetric to the theory of aggregate demand and would in some sense be a more elegant theory. However, much empirical support exists for the positive relationship between aggregate supply and the price level in the short-run, which poses a problem for the symmetric-treatment theory. The sticky-wage hypothesis is one mechanism by which an upward-sloping aggregate supply curve can be generated in this model of the macroeconomy. There are other mechanisms that have been developed to bridge this gap between theory and facts – however, none has yet proved to be completely satisfactory at both a theoretical and an empirical level.[51]

[51] Another important mechanism relies on firms' perceptions of the price of the goods they sell vis-à-vis the aggregate price level of the economy. Briefly, this line of thinking advances the argument that individual firms directly and immediately observe the price of the goods they sell, but do not immediately observe the overall price level of the economy. Firms only learn about the overall price level with some time lag, i.e., when news of inflation in the economy finally becomes available (after a few weeks or months, say). In the intervening time, firms may believe that the price of *only* their output good has increased, rather than all prices in the economy, which induces them to increase their output. Only when firms learn that all prices in the economy have changed do they return to their normal level of production. This is a more advanced theory – interested readers are referred to Abel & Bernanke, *Macroeconomics, 4th ed.*, Chapter 10 for a more detailed (yet readable) treatment of this version of the theory of short-run aggregate supply.

9. General Equilibrium

Equilibrium in the macroeconomy is determined by the interaction of aggregate demand and aggregate supply. Because we have distinguished between long-run aggregate supply and short-run aggregate supply, we also must distinguish between long-run equilibrium and short-run equilibrium. Loosely speaking, we will think of both of these kinds of equilibria as general equilibria. A **general equilibrium** occurs when *all* markets in the economy clear. To be completely precise, general equilibrium occurs in the model we have developed only when aggregate demand equals *long-run* aggregate supply. However, we will speak informally of a "short-run general equilibrium" as occurring when aggregate demand equals short-run aggregate supply. The distinction arises because of the different states of the labor market in short-run general equilibrium and long-general run equilibrium. Because of sticky wages, in a short-run general equilibrium, the labor market may *not* be in equilibrium. In a long-run general equilibrium, however, the labor market must be in equilibrium and hence *every* market in the economy is in equilibrium – which is the true sense of the term *general equilibrium.* Thus, a short-run general equilibrium occurs at point E in Figure 25.

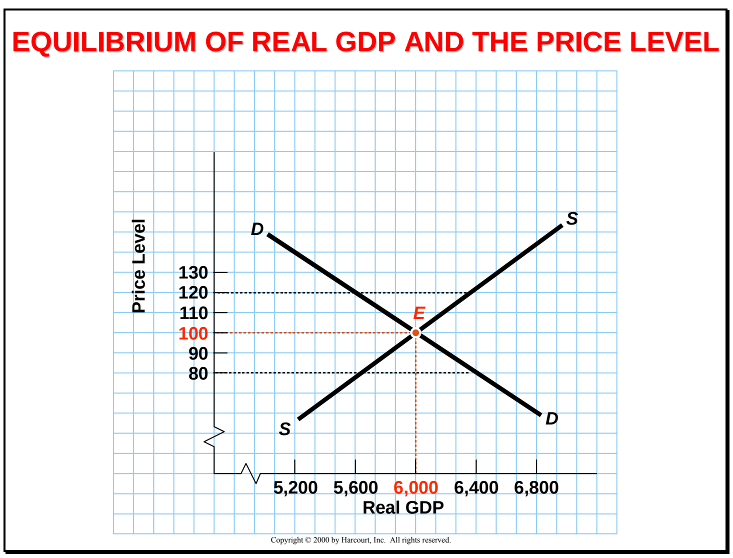

Figure 25

You should think of the mechanism depicted in Figure 25 in the same way as a supply-demand diagram from microeconomics. In particular, if aggregate demand is ever larger than aggregate supply, the price level will rise to re-equilibrate the system, and if aggregate demand is ever smaller than aggregate supply, the price level will fall to re-equilibrate the system. The point E shows the equilibrium price level and real GDP, (P^*, Y^*), that prevail in the economy. With the aid of diagrams like Figure 25, we can reconsider two concepts that were briefly introduced earlier – recessionary gaps and inflationary gaps.

9.1. Recessionary Gap

A **recessionary gap** occurs when short-run equilibrium GDP is below potential GDP. A recessionary gap can be visualized as in Figure 26.

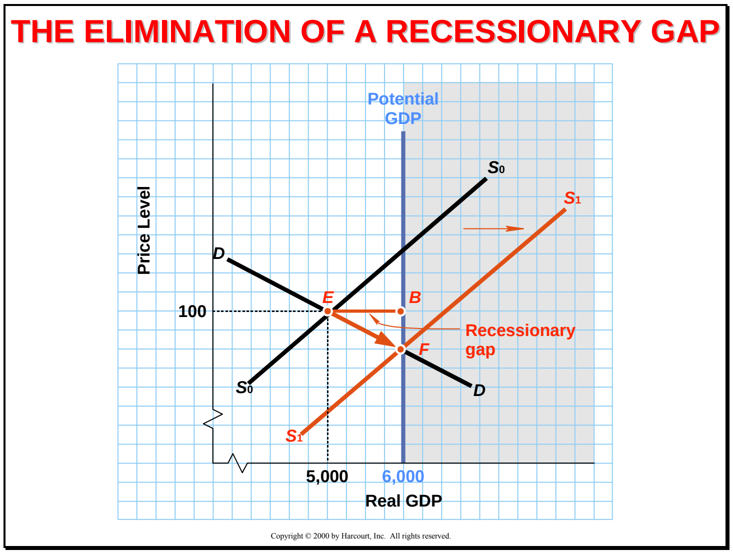

Figure 26

The mechanism by which a recessionary gap is eliminated is the following. In the presence of a recessionary gap, unemployment must be above its natural rate.[52] The relatively high level of unemployment will put downward pressure on wages because employers have more bargaining power than employees. When wages fall, the short-run aggregate supply curve will shift outwards,[53] reducing the recessionary gap. However, in the real economy, wages are downwardly sticky (because of wage contracts, certain regulations such as the minimum wage, and psychological resistance to nominal wage cuts), which may mean that a recessionary gap will be eliminated only with some time lag.

9.2. Inflationary Gap

An **inflationary gap** occurs when short-run equilibrium GDP is above potential GDP. An inflationary gap can be visualized as in Figure 27.

[52] Can you explain why?
[53] Recall the discussion from the section on aggregate supply.

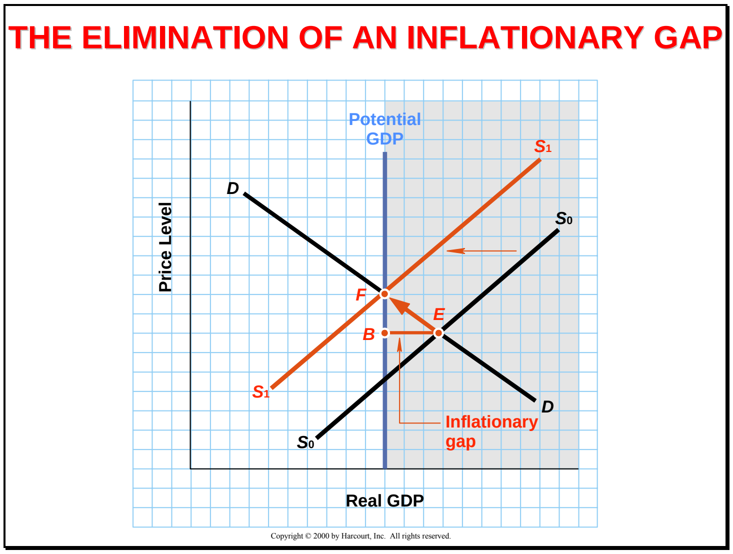

Figure 27

The mechanism by which the economy returns to a long-run general equilibrium in this case is the following. In the presence of an inflationary gap, unemployment must be below its natural rate. The relatively low level of unemployment will put upward pressure on wages because employees have more bargaining power than employers. When wages rise, the short-run aggregate supply curve will shift inwards, eliminating the inflationary gap. Notice that as the short-run aggregate supply curve shifts inwards to re-establish long-run equilibrium, the price level rises – hence the terminology *inflationary gap*. Also notice that during this transition, *prices are rising while output is falling* – this kind of "correction" of the economy is termed **stagflation** ("stag" for stagnant output growth and "flation" from inflation).

9.3. Did the U.S. enter a "New Economy" during the 1990s?

Many commentators have reported that the usual economic laws did not apply to the U.S. economy during the mid- and late-1990s because of the unprecedented combination of very low inflation and very low unemployment.

But the aggregate demand-aggregate supply analysis presented above can fully explain the economic events in the U.S. during this time. Using this framework of analysis, the experience of the U.S. economy can be interpreted very simply. The aggregate supply curve (*both* the long-run and the short-run supply curves) shifted out at rates faster than usual – which simultaneously led to faster output growth (and hence low unemployment) *and* low inflation. So the only "new" aspect here may the *reasons* why the aggregate supply curves shifted outwards more rapidly than usual – but by no means was it the case that the usual macroeconomic framework no longer applied. Indeed, the basic framework we have built can describe the experience of the U.S. economy during the 1990's quite well.

Possible reasons for the aggregate supply curves shifting out at such rapid rates in the past decade are rapid advances in information technology (which makes workers more productive), which is captured by rapid TFP growth. Recall from our study of aggregate supply that TFP growth shift both the short-run and long-run aggregate supply curves outwards.

9.4. Government Stabilization Policy

Because the economy's self-correcting mechanism sometimes works slowly (for example, the sticky wages which protract the elimination of recessionary gaps), it seems likely that government intervention can sometimes smooth out short-term disturbances in the economy.

As a simple example, consider the case of an aggregate demand curve that is deemed to be "too far to the left," so that short-run equilibrium output is too low relative to potential output. Because government spending is a component of aggregate demand, however, the government can increase its spending (that is, increase G), which will shift the aggregate demand curve to the right,[54] resulting in a higher equilibrium output. Alternatively, the government can decrease taxes, which will also shift the aggregate demand curve to the right. Finally, the Federal Reserve, through its control of interest rates, also has the power to influence aggregate demand. We will soon turn to a study of how such government policy affects macroeconomic outcomes.

[54] Recall the multiplier effect which drives the magnitude of the shift in the aggregate demand curve.

10. Fiscal Policy

The two broad categories of policy interventions that a government can undertake in order to stabilize an economy are fiscal policy and monetary policy. Broadly, monetary policy involves attempts to use nominal interest rates and the money supply in order to implement some desired macroeconomic outcome. In the U.S., monetary policy is conducted by the Federal Reserve. Fiscal policy involves manipulating government spending and taxes in order to implement some desired macroeconomic outcome. In the U.S., fiscal policy is conducted jointly by legislative and executive branches of government (i.e., Congress and the White House). The study of fiscal policy is the subject of this section, after which we will turn to a detailed study of money and monetary policy.

10.1. Revenue Sources and Expenditure Categories for the Government

Various taxes provide revenue (income) to the federal government. The most important of these sources are personal income taxes, social insurance taxes (i.e., Social Security payroll taxes), and corporate profit taxes. For the most part, we will simply assume that taxes are levied in lump-sum fashion – that is, each individual's tax burden is completely independent of any decisions he or it makes. In reality, of course, almost all taxes are specified as a percentage of some choice variable, such as income or consumption. These latter kinds of taxes are non-lump-sum (aka distortionary) taxes. Many of the important results we will consider hold for both lump-sum and non-lump-sum taxes – but considering lump-sum taxes greatly simplifies much of the analysis. However, an important branch of policy thought, loosely-termed supply-side economics, relies on the distortionary nature of taxes and how fiscal policy can reduce such distortions. As such, our consideration of fiscal policy will involve the use of both lump-sum and distortionary taxes.

The main categories of expenditure for the federal government are transfer payments (i.e., Social Security benefits, unemployment benefits), purchases of goods and services, and interest payments on the federal debt.[55] **Transfer payments are NOT included as part of what we have thus far been calling G.** For our purposes, we will simply consider the G part of the national income identity as government purchases of goods and services.

10.2. Government Deficits and Government Debt

If the federal government collects more in taxes than it spends on goods and services in some particular time period, then the federal government is said to have had a **fiscal**

[55] At June 2003, the U.S. federal government debt stood at approximately \$6.5 trillion – over half of annual GDP. Almost all of this debt is held by (domestic and foreign) holders of U.S. Treasury bonds.

surplus during that period. That is, for a fiscal surplus to have been realized in some time period, it must have been that $T-G>0$ during that time period. If $T-G<0$ during some particular time period, then the federal government is said to have had a **fiscal deficit** during that period. Note that fiscal deficits and surpluses are flow variables – that is, they are only meaningful when attached to some particular time period.

Federal debt, however, is a stock variable.[56] In any time period that the government runs a fiscal deficit, it must be that the financing shortfall was met through borrowing. For example, in the year 2001, the U.S. government ran a fiscal deficit of approximately \$80 billion, which means its spending was \$80 billion higher than its tax revenues. This gap was filled by U.S. government borrowing. The primary way that the U.S. government borrows is by issuing Treasury bonds, which are purchased by individual as well as institutional investors. Thus, the U.S. government owed \$80 billion more to bondholders at the end of 2001 than it did at the beginning of 2001 – hence its debt level rose by exactly the amount of the deficit. More generally, the government debt at any time is simply the sum of all past fiscal deficits minus the sum of all past fiscal surpluses.[57]

10.3. Fiscal Policy Effects on Aggregate Demand

Because government spending is a component of aggregate expenditure, an increase in government spending raises the aggregate expenditure schedule, which leads to a higher level of demand-side equilibrium real GDP. This causes an outwards shift of the aggregate demand curve. The amount by which the aggregate demand curve shifts out depends on the size of the multiplier on government spending.[58]

The new short-run equilibrium of the economy occurs at the new intersection of the aggregate demand curve and the short-run aggregate supply curve. This new short-run equilibrium clearly occurs at a higher price level and a higher real GDP. Thus, the short-run effect of an increase in government spending is increased economic growth and increased inflation. In the long-run, however, wages will rise to catch up to the inflation rate,[59] causing an inwards-shift of the short-run aggregate supply curve. In the long-run, then, the increase in GDP disappears and the price level simply rises even further.

The same qualitative effect can be achieved by lowering taxes. A decrease in taxes also leads to an outwards-shift of the aggregate demand curve (because consumers' disposable income rises), leading to a higher short-run price level and real GDP. In the long-run, wages rise, causing the short-run aggregate supply curve to shift leftwards and GDP returns to its potential level.

[56] Indeed, debt is by definition a stock variable, regardless of whether we are considering the debt of the federal government, a single consumer, or any economic agent.

[57] Analogously, in any time period in which there is a fiscal surplus, the debt declines by the amount of the surplus. That is, a surplus allows the government to pay off some of its bondholders.

[58] At this point, this discussion should simply be review. Refer to the earlier discussion of the multiplier effect and aggregate demand for more details.

[59] Recall the sticky-wage hypothesis used in deriving the short-run aggregate supply curve.

Both of these mechanisms for raising GDP in the short-run, increasing government spending and lowering taxes, are termed **expansionary fiscal policy,** simply because they both have the effect of (at least temporarily) expanding real GDP. **Contractionary fiscal policy** is defined analagously – it is the use of decreased government spending and/or increased taxes to lower (contract) real GDP.[60]

It is important to note here that the qualitative mechanism by which changes in taxes affect aggregate demand do *not* depend on whether or not taxes are lump-sum. In either case, a decrease in taxes (for the case of non-lump-sum taxes, a decrease in the tax *rate*) causes the aggregate demand curve to shift outwards, and an increase in taxes (for the case of non-lump-sum taxes, an increase in the tax *rate*) causes the aggregate demand curve to shift inwards.

10.4. Supply-Side Fiscal Policy

A large group of economists emphasize the effects of fiscal policy on aggregate supply, rather than the effects on aggregate demand as described above. The main idea behind this school of thought is that distortionary taxes, primarily on labor income, limit individuals' incentives to work and therefore boost production of the economy. As a theoretical notion, this way of thinking has much appeal. After all, the higher is the tax rate, the less incentive there is for a person to work a given amount of hours, simply because he gets to keep a smaller share of his total earnings. In the extreme case of a 100% tax on labor income, there would be no incentive to work at all. With a 100% tax rate on labor income, *all* of an individual's labor earnings would be taken away as taxes, leaving the individual with nothing – which, as far as the individual is concerned, could have simply been attained by not working at all.[61]

10.4.1. Laffer Curve

The above example illustrates that, in the presence of only distortionary taxes and no lump-sum taxes, *two* tax rates are consistent with zero total tax revenues collected by the government – 0% and 100%. Intermediate tax rates will therefore yield strictly positive tax revenues – but the fact that tax revenues are zero at the two extreme possible tax rates implies that tax revenues must be increasing over some region of tax rates and decreasing over some region of tax rates. This important implication of the incentive effects of taxes is formalized in the **Laffer Curve,** which relates how tax revenues depend on the tax rate.

[60] You should trace out the effects of such contractionary policies for yourself.
[61] A 100% tax rate may strike you as absurd. However, there have been instances of such tax rates in history. See Robert Shiller's *The New Financial Order* for an account of 100% tax rates centuries ago in parts of Europe. From more modern times, federal income tax rates of approximately 90% for the highest income earners in the U.S. in the early 1900's were common.

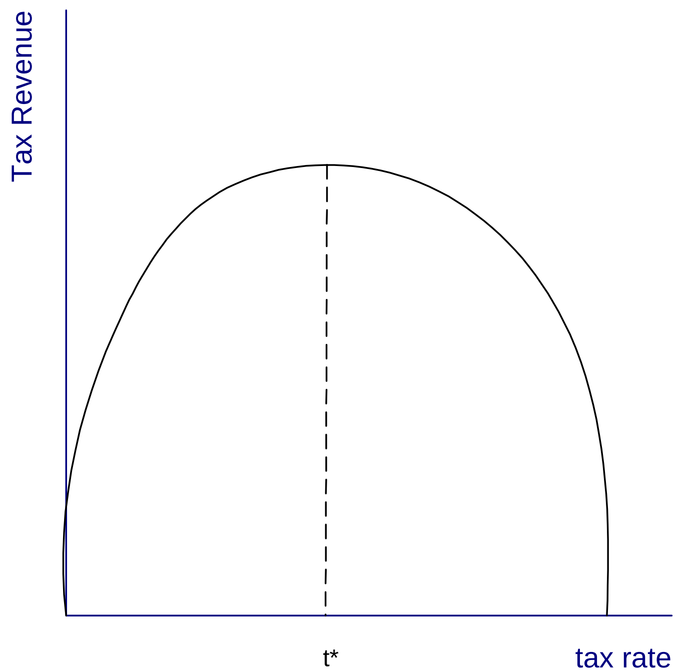

Figure 28

The Laffer Curve in Figure 28 shows that at some tax rate t^*, tax revenues are at their maximum. At tax rates higher than t^*, tax revenues actually decline (even though the tax *rate* is higher) – the reason is because of the negative impact that the high tax rate has on incentives to work and produce.

10.4.2. Theory Behind Supply-Side Fiscal Policy

As the above example suggests, when making their labor supply decision, what individuals care about is the *after-tax* real wage rate. Thus, the labor supply curve is a function $L^s\left(\frac{(1-t)W}{P}\right)$, where, as before, W is the nominal (money) wage rate, P is the price level, and t is the tax rate on labor income.[62] The labor supply curve is of course still upward sloping.[63] The introduction of a non-lump-sum labor income tax, assuming a fixed labor demand curve, causes a *non-parallel shift* of the labor supply curve, as illustrated in Figure 29. The reason that the shift is non-parallel is that the effect of a reduction in the labor tax rate is likely to be strongest on higher-paid workers because they pay more taxes than lower-paid workers.

[62] So t is a number between zero and one.

[63] Ignoring, as we did earlier, the conflicting income effects and substitution effects on labor supply. That is, we abstract away from the "backward-bending" labor supply curve.

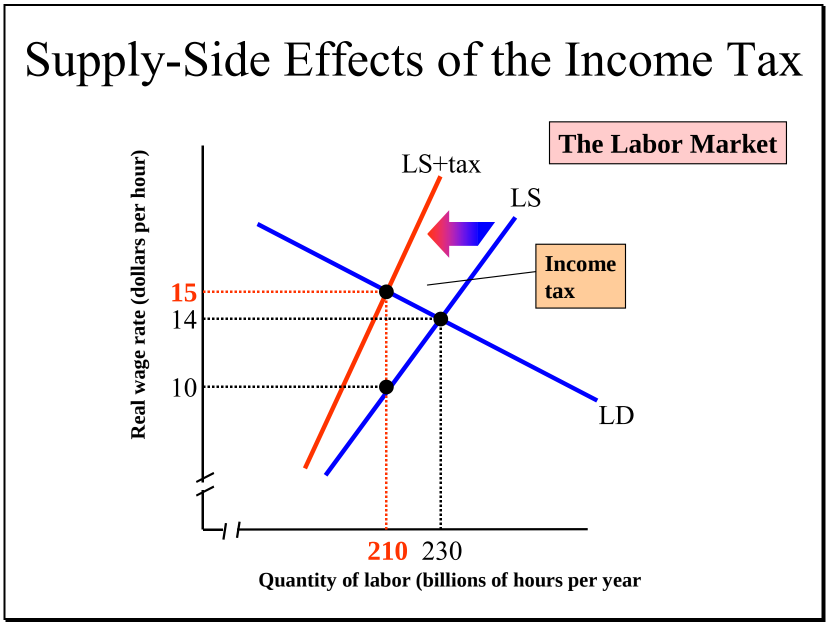

Figure 29

The effect on the equilibrium amount of labor hired in the labor market is larger than if the shift had been of the same amount at higher wages as it is at low wages. The end result is that a portion of the burden of the wage tax falls on workers and a portion also falls on firms (because notice that the *before-tax* wage rate W/P has risen from what it was without the wage tax – this burden is paid by the firms).[64] Thus, reducing the tax rate on labor income would cause exactly the reverse non-parallel shift of the labor supply curve.

Supply-siders believe that such a rise in labor, induced by a reduction of tax rates, will lead to a large outward shift of both the long-run and short-run run aggregate supply curves. The simple reason for the large outward shift of the short-run aggregate supply curve is that the aggregate production function shows that a larger amount of labor leads to increased production. The reason for the shift outwards of the long-run aggregate supply curve is that the drag on the economy caused by misaligned economic incentives manifests itself as a depressed value of total factory productivity, which we have been denoting A.

The main bite of the supply-side view is the belief that this outward shift of aggregate supply will be roughly of the same magnitude as the outward shift of the aggregate demand curve. This view is illustrated in Figure 30.

[64] Recall from basic microeconomics that the burden of a non-lump-sum tax (which might have been called an ad-valorem tax in microeconomics) in general falls on *both* demanders and suppliers, regardless of which side of the market is actually levied the tax. The relative burden between demanders and suppliers depends on the elasticities of the supply and demand curves.

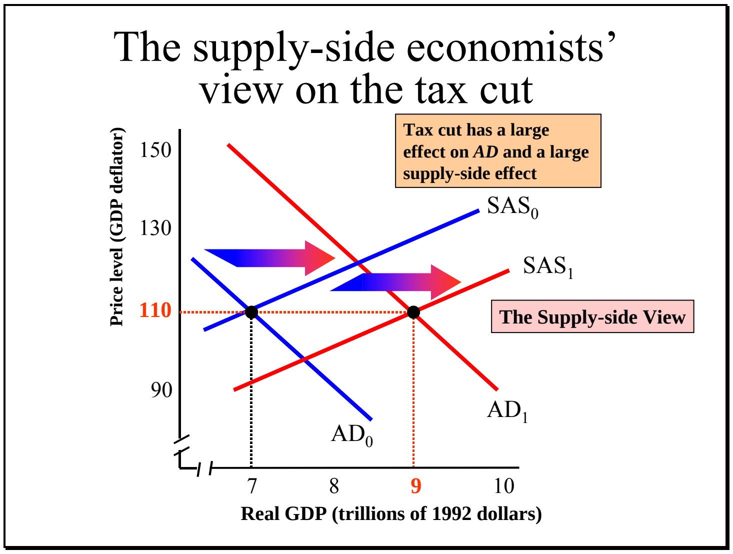

Figure 30

Figure 30 makes clear that supply-siders believe that the price level will change very little, if at all, in response to a tax cut. Indeed, the rise in real GDP is much larger than that which occurs if only the aggregate demand curve shifts out. It is this ultimate implication of supply-side effects that makes many economists in this camp quite passionate about their position.

10.4.3. Traditional View of Supply-Side Effects

Most economists have come to agree that the effects on labor supply and hence on aggregate supply of tax cuts are non-negligible. The main area of disagreement, then, is over the size of the supply-side effect. Figure 30 showed the supply-siders' view of the size of the supply-side effects of tax cuts. Figure 31 shows the more traditional view of the size of the supply-side effects of tax cuts.

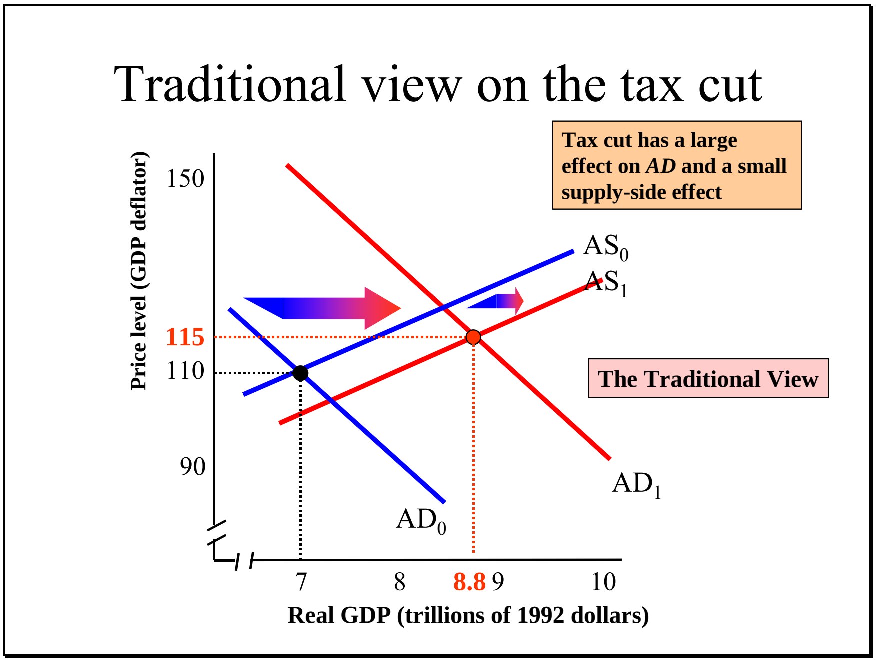

Figure 31

In this latter view, supply-side effects are acknowledged – however, the magnitude of the shift of the aggregate supply curve is much smaller than that of the aggregate demand curve. Thus, the price level still rises due to tax cuts, although by less than if there were no supply-side effects at all. The empirical importance of supply-side effects is an important unresolved issue in the study of macroeconomic policy.

10.5. Effects of Fiscal Policy on Interest Rates and Investment

One important effect that many economists emphasize is the effect of fiscal policy on real interest rates. We begin a brief consideration of this topic by first performing some algebraic manipulations starting with the basic national income identity. Ignoring net exports for now, the basic identity is $Y = C + I + G$. This expression can be rearranged to give

$$Y - C - G = I$$

The left-hand side of this last expression is called **national savings** and will be denoted by S^{nat}. The previous expression can also be written as

$$Y - C - T + T - G = I,$$

where lump-sum taxes T have simply been added and subtracted from the left-hand-side, obviously preserving the equality. We now define two more terms: **private savings**, denoted by $S^{priv} = Y - C - T$, is the total savings of all consumers in the economy, and

government savings, denoted by $S^{gov} = T - G$, is the total savings of the government. Notice that positive government savings is simply a fiscal surplus and that negative government savings is simply a fiscal deficit. Finally, it should be clear from these definitions that national savings is the sum of private savings and government savings – that is, $S^{nat} = S^{priv} + S^{gov}$.

We have already considered how investment, I, depends on the real interest rate r - investment is a decreasing function of the real interest rate. The effects of the real interest rate on national savings, if any exist at all, must occur through consumption. That is, in order to determine how national savings depends on the real interest rate, we must determine how consumption depends on the real interest rate.

It turns out that, theoretically, the way in which consumption depends on the real interest rate cannot be determined.[65] Thus, we must turn to empirical evidence on this matter. It turns out that empirical evidence does not help us much in making a determination. Most evidence suggests that the real interest rate has little effect on aggregate consumption – at best, a weak negative effect. That is, there is some evidence that the real interest rate and aggregate consumption may be *weakly* negatively correlated. For the purpose of the following discussion, we will assume that this negative correlation between the real interest rate and aggregate consumption is true. With this assumption, national savings must be positively correlated with the real interest rate, simply by the definition of national savings. This positive correlation between the real interest rate and national savings is illustrated in Figure 32, which shows an upward-sloping **savings schedule** along with the downward-sloping investment schedule.

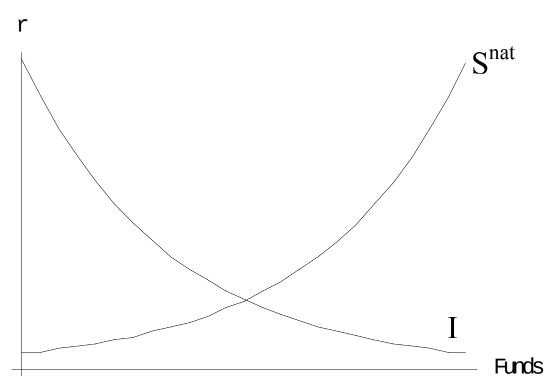

Figure 32

This figure looks very much like a market with some supply curve and some demand curve. Indeed, Figure 32 does illustrate a market – the **market for domestic funds.** The equilibrium real interest in the economy is given by the intersection of the national savings and investment curves.

[65] Recall that we mentioned this point when we first studied consumption.

With these tools and definitions, we can now consider the effect of fiscal policy on real interest rates.

10.5.1. Effects of Change in Government Spending

Consider first a change in government spending. For concreteness, assume that the government raises government spending, an expansionary policy. The definition of government savings, $S^{gov} = T - G$, shows that government savings declines – another equivalent way of stating this is simply that the fiscal surplus falls (because taxes are assumed to be held constant). The definition of national savings shows that national savings falls in response to the rise in government spending. In Figure 32, then, the entire national savings curve shifts to the left (because government spending is does not appear on the axes of that diagram). With the investment schedule unaffected, the equilibrium real interest rate in the economy has risen. The effect of this higher real interest rate is that investment declines (simply because of the negative relationship between the real interest rate and investment). Thus, a rise in government spending has led to a decline in investment by private firms – this important effect is referred to as **crowding-out.**

The opposite effect, that investment would increase if government spending decreases, is thus sometimes referred to as the **reverse-crowding out,** or the **crowding-in** effect. Crowding-out and crowding-in effects are usually important issues in fiscal policy deliberations.

10.5.2. Effects of Change in Lump-Sum Taxes – Ricardian Equivalence

Consider instead the effects on the real interest rate of a change in *lump-sum taxes*.[66] The definition of national savings shows that national savings is unchanged when lump-sum taxes change – simply because taxes do not appear at all in the definition of national savings! Taxes do appear, of course, in the definitions of private savings and government savings. It should be clear from the definitions that a change in lump-sum taxes affects private savings and government savings by exactly the same magnitude, but in opposite directions, which is why national savings is unaffected.

Because national savings is unaffected, the savings curve does not shift at all in response to a change in lump-sum taxes. Thus, the equilibrium real interest rate is unaffected. These important results are summarized by the **Ricardian Equivalence Proposition:** a change in lump-sum taxes, holding government spending constant, does not affect national savings and hence does not affect the equilibrium real interest rate.

The economic intuition behind Ricardian Equivalence is illustrated by the following example. Assume the federal government budget is currently balanced, i.e., $T = G$, but for some reason the federal government wants to cut taxes in the current period, while maintaining its level of government spending. This will clearly cause a fiscal shortfall in

[66] The main result to be described here, Ricardian Equivalence, holds *only* if taxes are lump-sum.

the current period. The government has only one option to cover this shortfall: it can issue bonds to the public (which are basically loans which the public makes to the government) in order to finance its spending. Now, the government is clearly running a fiscal deficit in the current period, i.e., $T - G < 0$, and its debt obligations have risen.[67]

Households now own pieces of paper (the bonds), which they know they will be able to redeem for the principal plus interest in the future (at the date of maturity of the bonds). The question to answer is whether households consider these bonds to be "wealth." If they do, then by the permanent income hypothesis, which says that consumption increases when wealth increases, consumption should rise due to the presence of these government bonds in the household's portfolio. However, fully rational households, assuming perfect information exists, will understand that these bonds do not represent increased wealth.

Harvard University economist Robert Barro[68] formalized the important insight that households will realize that future taxes will have to increase in order for the government to pay back the extra principal plus interest on the bonds that it issues to finance the current budget shortfall. In anticipation of the future tax increases, then, households will simply save the entire tax cut which they receive in the current period – that is, they will not increase their consumption. If consumption does not change, then neither does national savings, and the equilibrium real interest rate is unaffected.

10.5.3. Empirical Evidence on Ricardian Equivalence

Examining the time series movements of private savings ($S^{priv} = Y - C - T$) and government savings ($S^{gov} = T - G$) shows that they do broadly move in opposite directions – which implies that national savings is roughly constant over time. This then provides some empirical evidence for Ricardian Equivalence.

Of course, in reality, Ricardian Equivalence does not hold exactly. The major reasons why the result does not hold exactly in reality are:

1. Myopic consumers: Irrational individuals simply may not realize or understand that taxes will eventually have to be increased.
2. Finite lives: the individuals who pay the higher taxes in the future may not be the same as those who benefit from the lower taxes today (for example, some of those individuals might believe they will die before the higher future taxes are put into effect).
3. Liquidity-constrained consumers/imperfect financial markets: It may be the case that some individuals are not able to consume as much as they would like to because they are not permitted to borrow. In this case, lowering taxes will induce them to spend more today (a more thorough explanation of this would require a

[67] Recall that the federal debt is the sum of all past fiscal deficits minus the sum of all past fiscal surpluses.
[68] Barro, Robert, "Are Government Bonds Net Wealth?" *Journal of Political Economy,* 1974, volume 82, pages 1095-1117.

more complete analysis using indifference curves and budget constraints, which are more advanced topics).

Finally, if taxes are *not* lump-sum, Ricardian Equivalence does not hold.

10.6. Limits on the Effectiveness of Fiscal Policy

Note that all of the above assumes tacitly that the government accurately knows much about the economy. But in reality, the economic (and political) environment is constantly changing, which makes economic stabilization a difficult task at best. Even in the best of economic and political environments, economic variables change, policy targets move, the precise size of the multiplier is unknown, forecasts are uncertain, and fiscal policies take time to enact. For all these reasons, fiscal policy stabilization cannot be expected to produce exact results – it is the duty of responsible economics, however, to inform policy decisions with the best possible data and analysis.

11. Money and the Money Market

Before we can turn to a consideration of the effects of monetary policy on the macroeconomy, an introduction to the nature of money as well as an introduction to the structure of financial intermediaries is needed. In addition, we will briefly consider the nature of the banking system in most industrialized economies as well as the nature of banks' portfolios. Only then can the role, strengths, and shortcomings of monetary policy be analyzed.

11.1. The Role of Money

In the absence of "money," a *double coincidence of wants* would have to occur in any meeting between two economic agents for a possibility of trade to exist. A **double coincidence of wants** is said to occur when, in a chance meeting between two agents (call them agent 1 and agent 2), agent 1 wants the good that agents 2 has, *and* agent 2 wants the good that agent 1 has. If such a meeting occurs, the two agents will gain from trade (barter) with each other.

Clearly, in a modern economy, such a double coincidence of wants in a random meeting between two economic agents will happen very rarely. Thus, a system of barter, which involves the exchange of goods directly for other goods, would be extremely inefficient.

However, if there existed some commonly accepted *medium of exchange,* the frictions associated with barter exchange would be eliminated. Thus, "money" is best characterized as any object that serves as a medium of exchange.

Money is commonly thought to serve three functions in an economy:

1. Medium of exchange
2. Unit of account (an agreed measure for stating the prices of goods and services)
3. Store of value (although this feature is sometimes said to *not* be one of the defining characteristics of money)

11.2. Measuring Money

Defining what constitutes "money" is somewhat arbitrary, because essentially all that is required of a money is that it be accepted as a form of payment. It is possible for just about any object to be used as money – the fundamental difference between the possible forms of money is in their liquidities. **Liquidity** is the property of an object to be instantly converted into a means of payment with little loss in value

Despite the difficulty in properly defining money, some distinctions must be made – official government statistics define the following as major measures of the money supply in the U.S.:

> **M1 –** coins and paper money in circulation, plus checkable deposits (1998 value of $1.1 trillion)
> **M2 –** M1 plus money market deposit accounts, money market mutual funds, and savings accounts (1998 value of $4.2 trillion)

Until the early 1980's the monetary aggregate measures (M1, M2, and M3) used to be closely followed by the Federal Reserve in its monitoring of the condition of the economy. Indeed, the monetary aggregates used to be the prime channels through which the Federal Reserve attempted to implement its monetary policy objectives. Since then, however, the monetary aggregates are not paid nearly as much attention due to many reasons, one of the main being the widespread emergence of alternative means of payment, primarily credit cards. Note, however, that credit cards are *not* counted as part of the money supply.[69]

11.3. Financial Intermediaries

Financial intermediaries are firms that take deposits from households and firms and make loans to other households and firms. There are four major types of financial intermediaries:

- Commercial banks
- Savings and loan associations
- Savings banks and credit associations
- Money market mutual funds

These financial intermediaries are in fact the institutions that coordinate the activity in the market for domestic funds, which we were introduced to earlier. Simply put, financial intermediaries channel funds that savers supply to the funds market to the demanders of funds. Demanders of funds are primarily firms, and the demand for funds arises from desired investment spending.[70] We will focus below on the role of commercial banks in the transmission of monetary policy to GDP.

11.3.1. Balance Sheet of Commercial Banks

All banks have both assets and liabilities. The balance sheet of a bank lists its assets, liabilities, and net worth:

[69] Indeed, credit card usage, while providing a source of liquidity, is in fact a type of borrowing.
[70] Recall the discussion of the interaction of national savings and investment.

Assets – what the bank owns
Liabilities – what the bank owes
Net worth – the difference between assets and liabilities

A fundamental concept from accounting is:

$$Assets = Liabilities + Net\,worth$$

This accounting identity is *always* true (because of the *definition* of net worth).

11.3.2. Fractional Reserve Banking System

Banks earn profits by lending the money that customers deposit with it to other customers. This lending is possible because of the **fractional reserve banking system**, which is the system of banking that exists in the U.S. Under this system, a bank is legally required to only keep a fraction (call this fraction m) of its deposits on hand, ready to be returned to depositors on demand –the bank is free to lend out the rest if it wishes. The structure of the fractional reserve banking system leads to money creation by banks and is the key facet of the classical view of monetary policy transmission that will be discussed in detail below.

11.3.3. The Money Multiplier and Money Creation by Banks

Because a bank is only required to keep a fraction of its deposits on hand as reserves, it is free to loan the remainder to other firms and households. For the remainder of this section, assume the following:

- Banks keep only the required reserves (i.e., they keep zero excess reserves; this assumption can be justified because holding excess reserves entails an opportunity cost for banks because the excess reserves are funds that banks could lend and earn interest on);
- All new money created is held by individuals in the form of deposits, not currency (i.e., people do not put cash under their mattress, etc.) – this assumption is sometimes referred to as the **zero-currency drain** assumption.

Under these assumptions, when a cash deposit of, say, $\$D$ is received from a customer, a bank is able to loan $\$(1-m)D$ to some other customer (because it only keeps, due to the required reserve ratio, $\$mD$ as reserves).

The $\$(1-m)D$ loaned to this customer is deposited in some other bank, which is then obliged to keep a fraction m of *it* as reserves – which means that this second bank is free to loan $\$(1-m)^2 D$ to another customer.

This $\$(1-m)^2 D$ is then deposited in yet another bank – which, after meeting the required reserve ratio, is able to loan $\$(1-m)^3 D$ to yet another customer.

This chain continues indefinitely. Thus, the *total amount of new deposits* is given by the following infinite series:

$$\$D + \$(1-m)D + \$(1-m)^2 D + \$(1-m)^3 D + \ldots$$

which can be written in sigma-notation from mathematics as

$$\sum_{i=0}^{\infty} \$(1-m)^i D .$$

An infinite series has a compact algebraic expression[71] – in this case, we have

$$\begin{aligned} \sum_{i=o}^{\infty} \$(1-m)^i D &= \frac{1}{1-(1-m)} \cdot \$D \\ &= \frac{\$D}{m} \end{aligned}$$

Thus, through this *multiplier process* (and under the above two assumptions), each additional dollar of cash deposited at a bank ultimately leads to $\$1/m$ in new deposits. The term $1/m$ is called the **money multiplier**. In this sense, then, banks *do* create money.

You should be able to verify for yourself that if either or both of the assumptions above fails (specifically, the assumptions that banks hold zero excess reserves and that consumers hold none of the new money in the form of cash) that the money multiplier would be *less* than $1/m$. Thus, $1/m$ is an oversimplified form of the money multiplier.

One of the most important lessons to take away from this discussion is the following: the money supply, under a fractional reserve banking system, such as in the U.S., in the economy is *jointly* determined by the central bank, commercial banks, and consumers – it is *not* simply chosen by the central bank acting alone.

Also, the money multiplier process works exactly as you would expect it to in reverse, when the money supply contracts rather than expands, because a customer makes a cash withdrawal.[72]

11.4. The Money Market

[71] You may recall this from a precalculus or calculus course.
[72] Trace out the effects for yourself.

As in any other market, equilibrium in the market for money requires the joint consideration of both demand for money and supply of money.

11.4.1. Demand for Money

A common formulation of the model of consumers' demand for money holdings is that money demand depends positively on the volume of transactions and negatively on the nominal interest rate, that is,

$$M^D = M^D\left(nomGDP, i\right)$$

where *nomGDP* is the nominal GDP (which is a good proxy for the volume of transactions) and i is the nominal interest rate. For simplicity, the nominal interest rate that is meant here is the nominal interest rate paid on bonds, *not* an interest rate which might be paid on checking deposits. While it is true that in reality, some forms of money (i.e., checking accounts, savings accounts) do pay interest, to keep the analysis simple, we will assume that the only alternative asset which can be held in a portfolio of assets is nominal bonds – the interest rate on these bonds, therefore, is the *opportunity cost* of holding money and is thus the relevant interest rate which enters the money demand decision. The expression above is the **money demand function**. Graphically, the money demand curve in a graph of i versus M^D gives

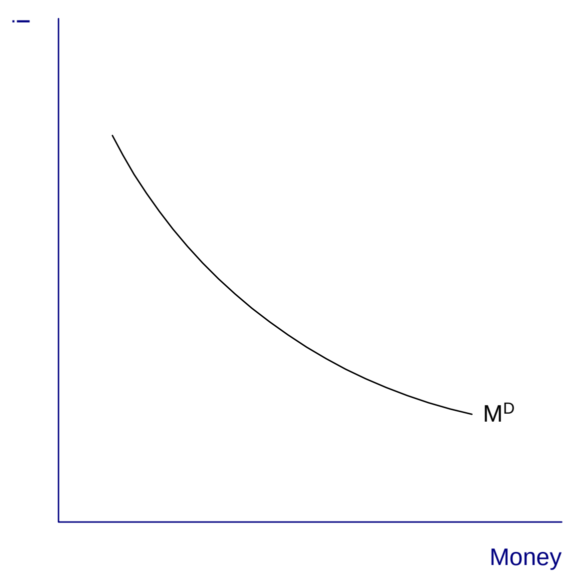

Figure 33

In this graph, note that nominal GDP, which we are using to represent the volume of transactions, is a shift factor of the money demand curve because it does not appear on the axes of this graph. For example, if nominal GDP rises, and all else remains the same, then the entire money demand schedule will shift to the right.

11.4.2. Supply of Money

Recall from above that the supply of money in the economy is jointly determined by the Fed (which decides the reserve ratio), commercial banks (which decide how much excess reserves to keep), and consumers (who decide how much of their money to keep as deposits and how much to keep as currency).

For any fixed Fed policy and any fixed "consumer policy" (that is, holding fixed consumers' decisions about what fraction of their money holdings to devote to currency), it is commercial banks who essentially determine the money supply. Because the nominal interest rate represents an opportunity cost of holding money for banks, the higher the nominal interest rate, the more loans banks will desire to make.[73] Thus, the money supply schedule determined in this way is an upward-sloping curve in a graph of i versus M^S. However, because this incentive for banks to expand their loan portfolios in response to a higher nominal interest rate is not crucial to the basic mechanism of monetary policy to be presented, we will abstract away from the upward-sloping money supply curve and simply represent the money supply of an economy as fixed for any given level of the nominal interest rate. That is, we will simply consider the case in which banks have made *all* loans that are profitable for them to make and that higher nominal interest rates do not increase their profit-maximizing loan quantity. Thus, the money supply schedule is a vertical line, as seen in Figure 34.

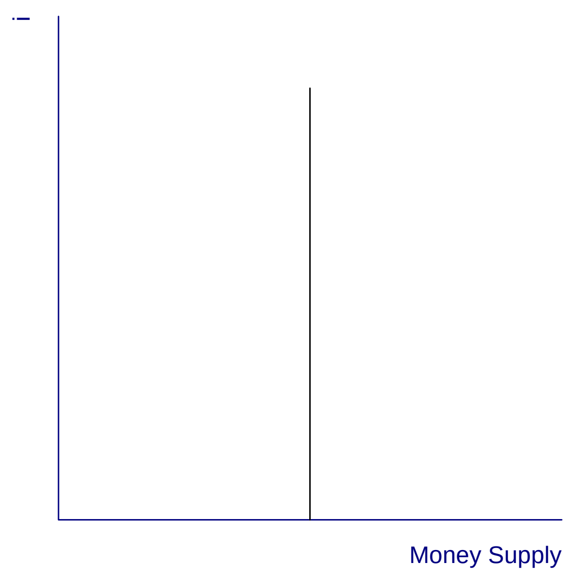

Figure 34

[73] Money that is simply sitting in a vault is not earning any interest *for the bank.* Thus, if the nominal interest rate is positive, banks will have an incentive to make loans. The higher the interest rate, the more loans banks will desire to make because of the higher potential profits – which will lead to a higher net worth of the bank. We abstract away from the upward-sloping money supply curve, however.

11.4.3. Money Market Equilibrium

It should not be surprising that money market equilibrium is the point where money demand and money supply are equal. Money market equilibrium is illustrated in Figure 35 as the point where the downward-sloping money demand schedule intersects the vertical money supply schedule.

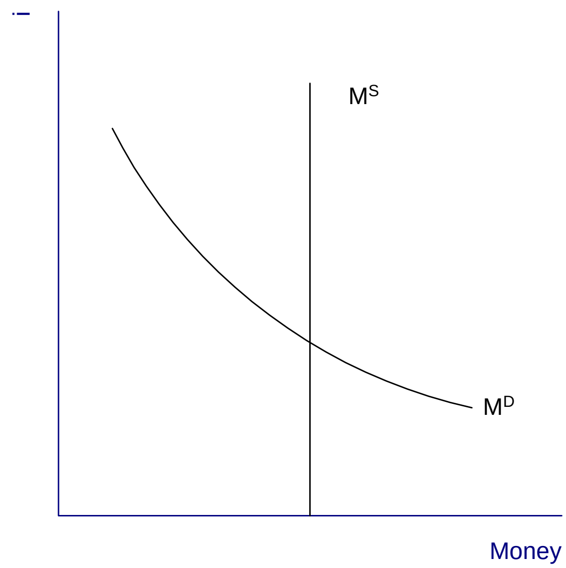

Figure 35

To make the analogy with standard microeconomic markets complete, the nominal interest rate is interpreted as the price of money. Thus, the above graphs are really simply graphs of the "price of money" versus the "quantity of money" – which you already are familiar with from basic microeconomics.

12. Monetary Policy

Now that we have introduced money, the financial system of the economy, and the money market, we are ready to begin a consideration of monetary policy, beginning with a discussion of the Federal Reserve System, the institution that is charged with implementing monetary policy in the U.S.

12.1. The Federal Reserve

The Federal Reserve System is the central bank of the U.S. – as such, it is often referred to as the "bank of the banks." The Fed (and a central bank in any country) is also often referred to as "the lender of last resort" – that is, the lender that banks themselves can turn to when they need loans.

The legislated goals of the Fed are:

1. Keep inflation in check (which means maintaining a relatively stable price level in the economy)
2. Maintain full unemployment
3. Moderate the effects of business cycles
4. Contribute towards long-run growth

Casual observation of these mandated goals of the Fed show that they are ambitious goals!

12.1.1. Structure of the Federal Reserve System

The Federal Reserve System is composed of more than just one bank. The system is composed of three major parts:

1. **Board of Governors –** the board is composed of seven members, each of whom is appointed to a 14-year term. The terms are staggered so that a vacancy arises every two years.
2. **Regional Banks –** there are 12 regional Federal Reserve Banks, each of which is charged with monitoring the conditions of its respective region of the U.S.
3. **Federal Open Market Committee –** this body meets approximately every six weeks to decide monetary policy issues. The meetings of this committee are those that are much-talked about in the media (i.e., "the Fed met today to decide interest rate policy")

12.1.2. Balance Sheet of the Federal Reserve

An example of the Fed's balance sheet would look like:

The Fed's Balance Sheet, December 1996

Assets (billions of dollars)		Liabilities (billions of dollars)	
Gold and foreign exchange	**21**	**Federal Reserve notes**	**427**
U.S. government securities	**460**	**Bank's deposits**	**25**
Loans to banks	**0**	**Other liabilities (net)**	**29**
Total assets	**481**	**Total liabilities (net)**	**481**

Figure 36

Notice that the Fed does not hold much gold in reserve – this is a consequence of the fact that the U.S. currency is not backed by gold, as it once used to be. That is, U.S. currency is not convertible into any other real asset, except as deemed by markets – the government makes no promise to its currency holders of what it can be converted into.[74]

Also notice the entry named "Bank Deposits" under liabilities. This entry reflects the fact that commercial banks place their deposits in accounts with the Fed.

12.1.3. The Fed's Instruments of Monetary Policy

We saw above that *private banks do create money* in the U.S. financial system – a power that is often incorrectly thought to be exclusively that of the Fed.[75]

Because of many financial innovations which have occurred in the U.S. over the past two decades (e.g., the increased use of credit cards, the ability to write checks against accounts such as brokerage accounts), M1 and M2 are not as closely watched by the Fed as they used to be, because they may not reflect as accurately "liquidity supply" in the economy today as before these various innovations in payments technologies. As a result of this, the Fed no longer targets the size of M1 and M2 as instruments of monetary

[74] This remains to this day a common misperception of the nature of U.S. currency – while all currency was once backed by gold deposits, the gold standard was officially abandoned after World War II. Thus, the only thing backing U.S. currency is the "full faith and credit" of the U.S. government. Such money is called **fiat money –** items (i.e., pieces of paper) which are money simply because the government declares them to be money.

[75] The quantity of money under direct control of the Fed is called the monetary base, or sometimes "high-powered money."

policy, as it did for a short while in the late 1970's. What, then, are the targets of monetary policy? There are three principal instruments used today:

1. **Open Market Operations:** Through "open market operations" (which are traditionally conducted by the New York Fed), the Fed can raise or lower the amount of reserves the banking system has. These new reserves are excess reserves for banks – and as we saw above, there will usually be multiple rounds of lending as a result of these new excess reserves, which raises the money supply in the economy in a manner consistent with the money multiplier. Open market operations are conducted via U.S. Treasury securities. Consider the case in which the Fed wishes to expand the money supply. The way the Fed can achieve this goal through open market operations is by buying Treasury securities from commercial banks (banks generally hold Treasury securities among the holdings of their assets). In the process of "buying" securities from commercial banks, the Fed gives money to the banks, and in exchange the banks give securities to the Fed. Thus, in the balance sheet of the Fed, both the items "U.S. government securities" and "Bank Deposits" would increase by the appropriate amount (see the above figure – and recall the basic accounting identity $Assets = Liabilities + NetWorth$).
2. **Required Reserve Ratios:** In practice, the Fed can alter the required reserve ratio (which, recall, we denoted above by m), which then affects the amount of potential money that can exist in the economy. (Note the use of the word "potential" – this is because of the critical assumptions we used in deriving the oversimplified formula for the money multiplier above, specifically zero excess reserves and zero currency drain.) However, the Fed rarely alters the required reserve ratio – it has been constant at 10% since 1992.
3. **Nominal Interest Rates:** The major way that the Fed implements monetary policy is by targeting interest rates. This is because interest rates still do reflect the "cost of money," regardless of how murky the definition of money itself may have become. Two of the main interest-rate levers[76] by which the Federal Reserve attempts to implement its monetary policy are:

   Discount rate – the interest rate the Fed charges to banks that need to borrow in cases when their required reserves temporarily fall below the required amount; this interest rate is *directly* controlled by the Fed.

   Federal Funds Rate – the interest rate that banks charge to other banks for loans to cover temporary shortfalls in required reserves; this interest rate is *market-determined* – that is, it is not under the direct control of the Fed, even though the Fed often sets "targets" for the Federal Funds Rate.

The discount rate is the interest rate that banks face when they borrow from the Fed directly, while the federal funds rate is the interest rate that banks face when they borrow

[76] Indeed, many would say these are the *only* interest-rate levers the Fed has at its disposal.

from each other. In reality, banks that do have temporary reserve shortfalls are reluctant to borrow from the Fed, because that would send a very bad signal to the customers of the bank. Thus, the discount rate is not usually as important a nominal interest rate in the U.S. economy as is the Federal Funds rate. The Federal Funds rate is the one that is usually referred to in media headlines such as "The Fed lowers interest rates again." Keep in mind, however, that the Fed actually does *not* set this interest rate directly, even if it may have a target value of the Federal Funds rate in mind.[77]

12.2. Transmission Mechanism of Monetary Policy

We now turn our attention to how the conduct of monetary policy by the Fed affects both the monetary side of the economy as well as the real side of the economy.

Consider the case of the Fed seeking to lower interest rates through an open market operation. The mechanism by which the goal would be achieved is:

> Fed buys bonds from banks (new money enters the economy) → money supply curve shifts out → money market equilibrium occurs at a new lower interest rate

The above schematic represents the case of *unanticipated* moves by the Fed. However, as should sound natural, economic agents may be able to anticipate moves by the Fed. In this latter case, current market conditions may already reflect the anticipated move.[78] Also note that the above mechanism is silent on what happens to the *real* side of the economy – that is, we have not said anything about how (or even if) monetary policy affects measures of (real) economic activity such as aggregate demand and GDP.

12.2.1. Classical Transmission Mechanism: how monetary policy affects the real side of the economy

In order to affect real components of the economy, a change in the nominal interest rate must lead to a change in the real interest rate. The reason why this can occur (and this can *only* occur in the short-run) is that inflation expectations are slow to adjust. If inflation expectations are slow to adjust, the Fisher equation shows that a decrease in the nominal interest rate leads to a decrease in the real interest rate.

[77] For example, between November 2001 and March 2002, the Fed's target for the Federal Funds rate was 1.75%. Following the actual Federal Funds rate, however, shows that it fluctuated around this target during this time period – for example, in early January 2002, it was at 1.69% and in mid-February 2002 it was at 1.77%.

[78] As an example from current events, consider the Fed's half-point reduction in the discount rate on March 20, 2001. The immediate effect on bond prices (as of the afternoon of March 20) was that they remained virtually unchanged. The main reason for this seemingly counterfactual result is that bond markets had already expected the Fed to take action and thus bond prices already reflected the expected rate cuts.

We have already seen that the component of GDP which is most affected by real interest rate is investment. Policy discussions of interest rate changes often also consider their potential effects on consumption of certain "big-ticket" items, such as cars and appliances. We will de-emphasize this latter channel, motivated mainly by the lack of theoretical support and the mixed empirical support for any link between real interest rates and consumption. Thus, the primary way in which monetary policy affects the real side of the economy is through investment. Schematically, monetary policy affects the real side of the economy in the following way:

> Lower real interest rate → the economy moves **down and to the right** along the investment curve → the rise in investment leads to a shift outwards of the aggregate demand curve → short-run GDP and short-run price level rises

Notice carefully the second item in the transmission mechanism: the investment curve does not shift outwards, but rather there is a movement along the investment curve. If the investment curve were shifting outward, that would imply a rise in the real interest rate of the economy (think of the savings-investment diagram), contrary to where we started this thought experiment which is that the Fed is lowering the nominal interest rate and hence the real rate.

12.2.2. Quantity Theory of Money

A complementary school of thought, the monetarist school, contends that the classical transmission mechanism described above works only faintly at best, if at all, because of the many uncertainties that prevail in modern economies and especially because the classical transmission mechanism makes a statement only about the short-run impact of monetary policy. The monetarist school of thought holds the view that, at least in the long-run, all changes in the money supply are reflected in long-run changes in the price level, and output in the long-run will be unaffected.

The monetarists' argument rests fundamentally on the following expression, which is often referred to as the **equation of exchange**:

$$PY = MV,$$

where P is the price level, Y is output, M is the quantity of money in the economy, and V is the **velocity of money.** The equation of exchange states that the entire nominal GDP (which after all is simply the sum of the money values of all final goods and services produced in some time period) of an economy must be *paid for* by the available quantity of money in the economy. The smaller the amount (stock) of money in the economy, the more times, on average, one unit of money must circulate through the economy.[79] Thus V represents the number of times, on average, that a unit of money circulates through the economy in some given length of time – the velocity of money.

Monetarists hypothesize that money demand is proportional to the volume of transactions (which, recall from above, can be proxied by nominal GDP):

$$M^D = k \cdot PY \,.$$

Here, k is just some constant of proportionality. Now, note from the equation of exchange, that $PY/M = V$. Thus, it must be the case that $k = 1/V$. So the constant of proportionality is simply the inverse of the velocity of money, in this specification of the money demand function. **Notice that in this monetarist money demand function, the nominal interest rate does not appear.**

Expressing the equation of exchange in terms of percent changes, we get

$$\%\Delta P + \%\Delta Y = \%\Delta M + \%\Delta V$$

which can be rewritten by means of one simple algebraic rearrangement as

$$\%\Delta P = \%\Delta M + \%\Delta V - \%\Delta Y \,.$$

Monetarists believe that the constant k is very stable over time – which implies that the velocity of money is very stable over time. Symbolically, this means that $\%\Delta V \approx 0$ (the percent change in money velocity over time is very near zero). This means, that any change in the amount of money must be reflected in either a change in output or a change in the price level. There is mixed empirical support for the monetarists' claim, however, of stable money velocity. Figure 37 shows that velocity of M2 money (Panel b) does appear quite stable over time, but the velocity of M1 money (Panel a) is not stable.

[79] And of course, the larger is the stock of money, the fewer times, on average, one unit of money must circulate through the economy.

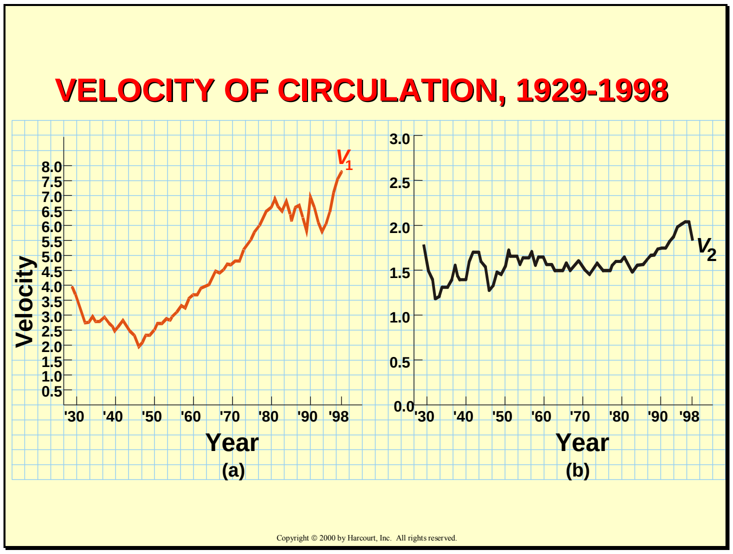

Figure 37. Panel a shows the velocity of M1 money, and panel b shows the velocity of M2 money over the period 1929-1998. Clearly, M2 velocity is more stable than M1 velocity.

In the long-run, the monetarist view is that only prices will be affected by changes in the money supply. Thus, monetarists favor a central bank policy which is based on a constant or stable money supply rule, rather than on "discretionary" monetary policy, in which the central bank, in essence, takes actions when "it seems like" it is the appropriate time to do so. The fundamental reasons that monetarists hold this view is that there are too many points of uncertainty in what is considered the classical transmission mechanism and because money velocity is constant over time.

Note, however, that the classical view of monetary policy transmission and the monetarist view are not at odds – the monetarist view simply emphasizes the long-run relationship between money and real variables, while the classical view emphasizes the short-run relationship between money and real variables. Thus, what is often portrayed as a great divide between these two views is largely a red herring.[80]

[80] To be even more precise, the monetarist view presumes that velocity is a *stable function* of the nominal interest rate, rather than literally constant over time. When one recognizes this view of monetarism, the distinction between it and the classical view becomes even less sharp.

13. Phillips Curve

The Phillips curve is a statistical relationship that relates the real side of the economy to the monetary side of the economy. To preview the main result**,** the Phillips relation simply states that in, in the short-run, the higher is the rate of inflation, the lower is the rate of unemployment. Graphically, the Phillips relation is illustrated in Figure 38.

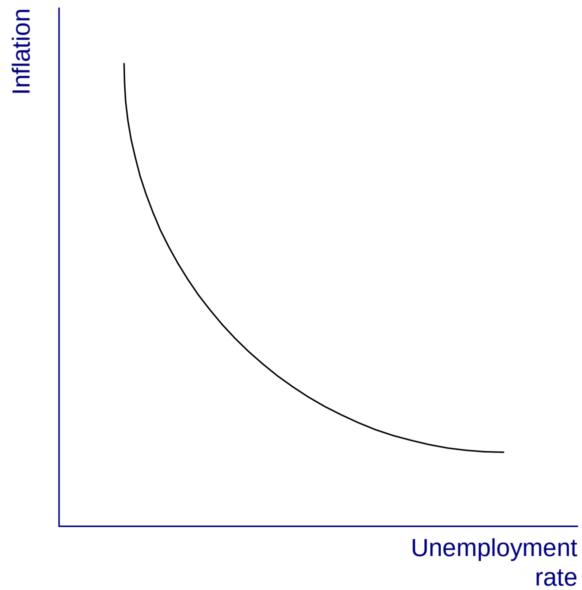

Figure 38

The main result depends on the assumption that changes in real GDP arise from fluctuations in the aggregate demand curve. That is, holding the short-run aggregate supply curve fixed, shifts in aggregate demand are the sole reason for changes in the price level and real GDP. But deviations of real GDP from potential GDP occur because of fluctuations in the labor market. Thus, another way of considering the effects of shifts in aggregate demand holding the short-run aggregate supply curve fixed is that inflation and unemployment are inversely related to each other. This latter relationship is exactly what the Philips curve describes.

On the other hand, if changes in real GDP arise from shifts in the short-run aggregate supply curve, then *higher* inflation would be associated with *higher* unemployment. Such a relationship is *not* what a given Phillips curve describes. However, in the context of the Philips relation, shifts of the short-run aggregate supply curve are associated with *shifts of the Philips curve.*

Another way to understand the Phillips relation is by first referring back to the underpinnings of labor demand. Recall from earlier the labor demand curve:

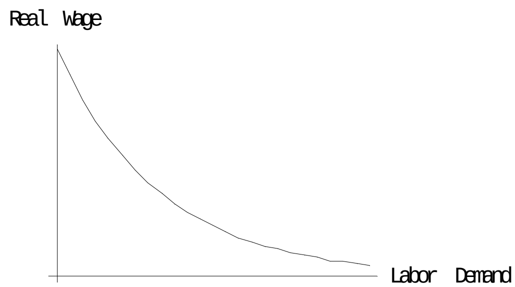

Figure 39

In many sectors of the economy, compensation rates for employees are fixed in advance for some period of time. For example, a union may have a contract that states that each worker will be paid $20 per hour for every hour members of the union work in the next one year. Presumably, this contracted wage represents some negotiation that occurred between the firm and the union. It seems reasonable to believe that this negotiation, and thus the contracted wage, reflects some anticipation about what the price level will be one year from the time the contract is signed.

Now consider what happens if the price level one year later actually turns out to be higher than expected (that is, that actual inflation turns out to be higher than expected inflation). In this case, the real wage W/P is *lower* than expected, which gives the firm an incentive to hire more workers than it was originally planning to hire. This means that unemployment will decrease – which also establishes the inverse relationship between inflation and the unemployment rate[81] that the Phillips curve embodies.

Note in the diagram above of the Phillips curve that *expected inflation* does not appear on the axes – which means that expected inflation is a shift factor of the Phillips curve. (It is *actual* inflation which appears on the axis of the diagram above.)

To explain different points of view of how the Phillips curve may be shaped and how inflation dynamics work, we first briefly consider different kinds of expectation formation mechanisms.

13.1. Expectations Formation

Economic models often need to make some assumption about *how* the agents in an economy develop their beliefs about their economic environment. There are three major expectation formation specifications that are commonly used in macroeconomic research:

1. **Perfect Foresight –** economic agents can perfectly forecast what all economic variables will be in the future (clearly this is an unrealistic

[81] Convince yourself that you can argue the result when the price level turns out to be lower than expected.

assumption but it often allows tractable analysis, and that is why it is sometimes used)

2. **Adaptive Expectations –** economic agents simply look at the past when they form their expectations about the future. For example, if the inflation rate in the country has been 10% for some time, people believe that the inflation rate in the future will continue to be 10%.
3. **Rational Expectations –** economic agents optimally use all available information (not simply past values of variables as in adaptive expectations above) in forming their expectations about the future. This means that agents will not systematically be incorrect in their expectations.

13.2. Adaptive Expectations and the Clockwise Loop of Inflation Dynamics

As mentioned above, expected inflation is a shift factor of the Phillips curve. If the government is implementing an inflationary policy (that is, one designed to raise the rate of inflation) *and agents have adaptive expectations,* then a "clockwise loop" of inflation dynamics results, which is stylistically presented in Figure 40.

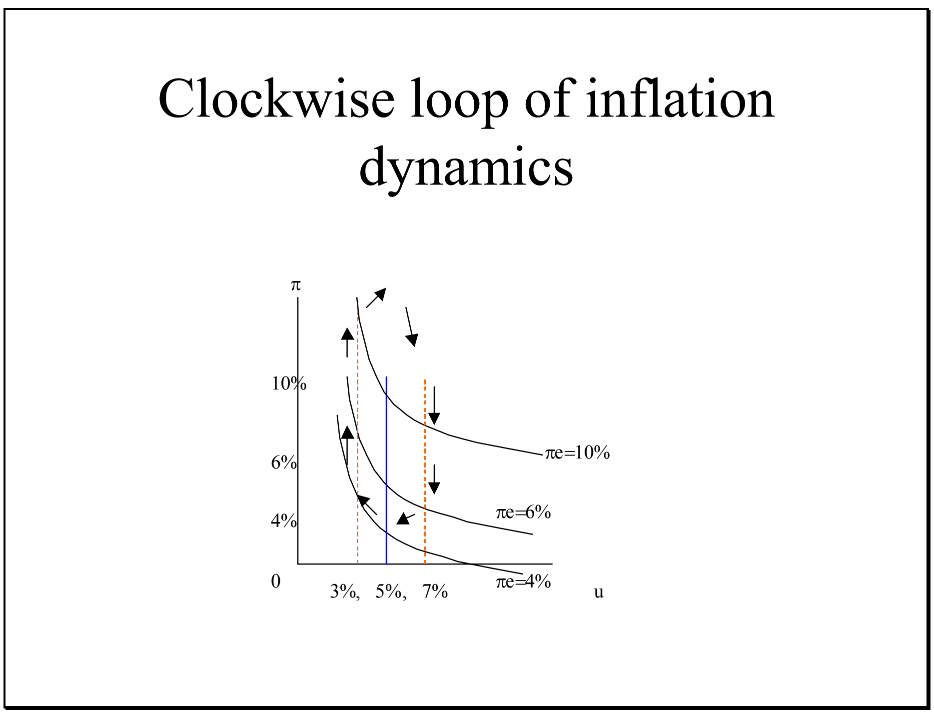

Figure 40

This pattern of dynamics emerges because agents' expectations of inflation is consistently lagging what the actual inflation rate is. That is, by the time agents have adjusted their expectations about inflation in accordance with actual past inflation, the actual inflation rate has changed. Agents' expectations are always "one step behind" what is actually occurring. Thus, by the time the Phillips curve has shifted, the inflation rate that the government is targeting has changed.

At some point, it is likely that the very high inflation rate becomes unsustainable, at which point a disinflationary[82] policy will be implemented. But, again, because agents have adaptive expectations, their expectations about inflation lag actual inflation. Thus, "on the way down the loop," the Phillips curve shifts inwards, but by the time each shift inwards occurs, the actual inflation rate has already fallen yet further. Tracing the path of inflation and unemployment from beginning to end yields the clockwise loop path.

13.3. Rational Expectations and the Vertical Long-Run Phillips Curve

If we assume that agents have rational expectations, then past values of inflation is but one set of information used in deciding what future inflation might be. Because (under the strong assumption that all agents have access to the same information) agents know that the government intends to pursue an inflationary policy, their expectations about future inflation will, on average, be correct. That is, there will be no lag in expected inflation vis-à-vis actual inflation. Thus, the Phillips curve will be vertical.

It seems reasonable, in any case, that the longer the time period under consideration, the more "rational" agents will be in forming their expectations. That is, even if agents can be "fooled" once or a few times, it is unlikely that they will be fooled over and over again in the long-run. Thus, it is likely that the long-run Phillips curve is vertical. This means that, in the long-run, there is *no* tradeoff between inflation and unemployment. Indeed, such a long-run relationship emerges even in the case of adaptive expectations considered above.

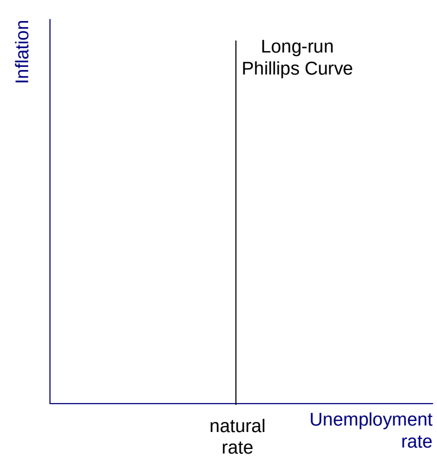

Figure 41

[82] Recall that disinflation occurs when the rate of inflation decreases – disinflation is *distinct* from deflation, which is a decline in the price level.

14. International Economics I: International Trade

The macroeconomic analyses we have conducted so far have been for the case of a closed economy, one that does not trade with the rest of the world. Formally, the assumption of a closed economy amounts to $NX = 0$. Strictly speaking, because $NX = X - IM$, $NX = 0$ does not literally mean that the economy is closed to international trade – it only means that exports and imports exactly balance each other. Regardless of which interpretation we adopt, the analysis as conducted so far is unaffected. But even casual observation of events points out that economies are increasingly interdependent, suggesting that a closed-economy analysis must miss important elements of actual macroeconomic phenomena. We now extend our macroeconomic model to incorporate an international sector, beginning first with an examination of international trade, continuing with an introduction to the international financial system, and then concluding with a modified analysis of economic policy that acknowledges that the international sector interacts with policy decisions to sometimes yield different macroeconomic outcomes than in the closed-economy analysis.

14.1. Absolute Advantage and Comparative Advantage

A natural question to ask is why *do* countries engage in trade? Under the belief that trade is voluntary (that is, countries are not *forced* to trade with other countries), the fact that trade does exist between any pair of countries indicates that both sides must expect to gain from trade.[83]

More formally, the underlying reason why gains from trade do exist for any pair of countries is the concept of comparative advantage. Two definitions are in order before we proceed:

Absolute advantage: a country is said to have absolute advantage over another in the production of a particular good if it can produce that good using smaller quantities of resources than can another country

Comparative advantage: a country is said to have comparative advantage over another in the production of a particular good if it produces that good *less inefficiently* than the other country.

Consider the example illustrated in Figure 42, which shows straight-line production possibilities frontiers (PPFs) for Japan and the U.S.[84] These PPFs show the trade-off in the production of television sets and computers in the two countries.

[83] Otherwise why would they trade?

[84] Recall from microeconomics that a straight-line PPF indicates that there is no specialization of resources.

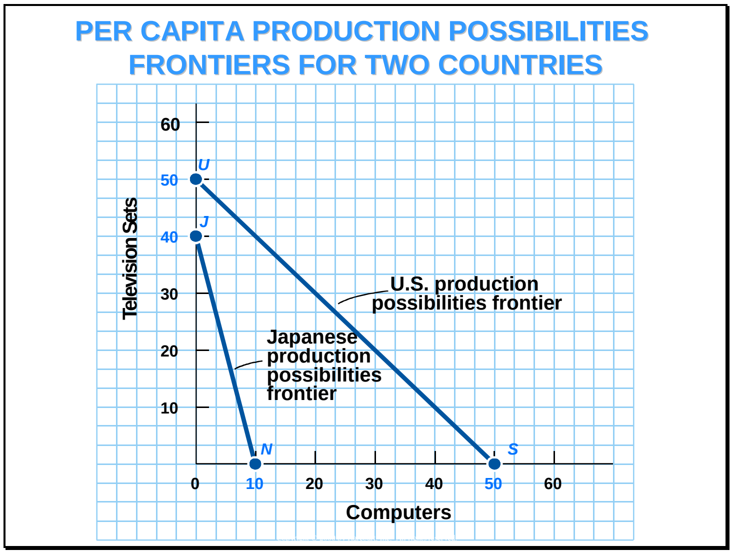

Figure 42

The fact that the line JN is completely inside the line US indicates that the U.S. enjoys absolute advantage in the production of the both TVs and computers. The fact that the slopes of the two lines are different indicates that each country enjoys a comparative advantage in one of the goods – in this case, Japan has a comparative advantage in TVs and the U.S. has a comparative advantage in computers.[85]

The slopes of the PPFs indicate the tradeoffs between the two goods in each country *if each country was in autarky.*[86] For Japan, this tradeoff is four TVs for one computer, and for the U.S. this tradeoff is one TV for one computer.

Now imagine that each country *completely specializes* its production in that good in which it has a comparative advantage. Also, for the purpose of illustration, say the rate of tradeoff between TVs and computers *in the world*[87] is two TVs for one computer. If Japan and the U.S. both trade freely with the rest of the world, then the *consumption* possibilities frontiers (CPFs) for each country, superimposed on the same graph as the PPFs, are as in Figure 43.

[85] Verify that this is correct using the definition of comparative advantage given above.
[86] **Autarky** – a country is said to be in autarky if it does *not* trade at all with the rest of the world.
[87] Another way to state this is that the *world* price of computers in terms of TVs is two. Notice this is just an example – nothing we have said thus far implies that the world price *must* be two TVs for one computer. All we can say at this point is that world price of computers in terms of TVs will be something in between one (the price in the U.S. in autarky) and four (the price in Japan in autarky).

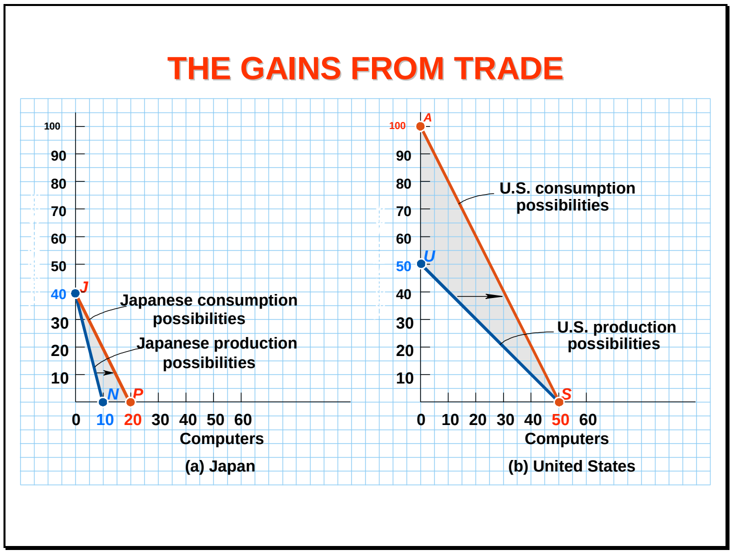

Figure 43

The fact that each country's CPF is outside its PPF shows that there *are* gains from trade for each country – simply because the set of consumption choices that the consumers of each country now have available to them are strictly larger than the set of choices they had available to them in autarky.

14.2. Determination of World Prices

The above example simply assumed a world price of computers in terms of TVs. As you may expect, the world price is actually determined by supply and demand, and the analysis can be performed using usual supply and demand curves – with a twist.

An important observation is that, for each good which is traded on the world markets, if a country is exporting a good, it must be that the world price of that good is *higher* than the equilibrium price which would prevail in autarky in that country. Otherwise, there would be an incentive for that country to *import* that good, rather than export it. Similarly, if a country is importing a good, it must be that the world price of that good is *lower* than the equilibrium price which would prevail in autarky. For simplicity, imagine that the whole world is made up of only two countries.[88] Then it must be that, for any good, the amount that one country exports must be equal to the amount that the other country imports. This last statement simply says that the *world market* for that good clears – that is, *world demand equals world supply*. Putting all these observations together, we see that the world price is determined graphically as in Figure 44.

[88] All of the results to be described apply for the case of more than two countries as well – the two-country case highlights the main point more clearly.

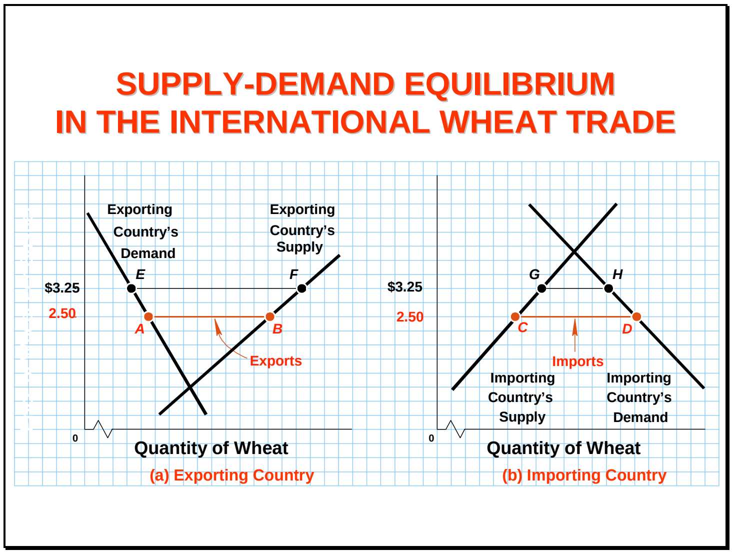

Figure 44

Figure 44 shows that the world equilibrium price is determined where world supply equals world demand – or, in other words, the world price is that price at which imports in one country are exactly equal to exports in the other country. In Figure 44, the line segment AB in the left panel is equal in length to the line segment CD in the right panel. The price at which the world market clears is the world price of that good – here, the world price is $2.50. This price is the only price at which the world market clears.[89]

14.3. Distributional Consequences of Trade

A popular argument advanced in favor of restricting international trade is that because labor costs are "cheap" in so many other parts of the world relative to the U.S., U.S. workers lose as a result of free trade. However, as we have seen earlier, wages are set such that real wages are equal to the marginal product of labor, that is

$$\frac{W}{P} = MP_L .$$

Under perfect competition, this condition is true in **every** country in the world. Thus, a low real wage signifies a low marginal product of labor – so "cheap" labor is a result of low productivity! It is hard to imagine that countries with much lower productivities can "take jobs away" from U.S workers. Thus, this kind of argument in favor of protectionism may not be a sound economic argument.

[89] To verify your understanding of how world prices of goods are determined, determine the world price of the good in Figure 44 if the autarky prices in each of the two countries are equal to each other. In this case, are there any gains to trade? What would the respective PPFs for the two countries look like in this case?

However, while there are gains from trade for each country as a whole (as shown in Figure 43), there certainly are distributional consequences for different groups of individuals within both the importing and exporting countries.

Simply examining Figure 44, we can see that in the exporting country, suppliers (of the good under consideration) benefit because the world price is higher than that country's autarky price, while consumers (of the good under consideration) are hurt for the same reason. In the importing country, the fact that the world price is lower than that country's autarky price means that consumers benefit while suppliers are hurt.

14.4. Barriers to Trade

Continuing to examine Figure 44, because the suppliers of the good in question in the importing country are being hurt by the open trade arrangement with the exporting country, they may try to organize themselves and lobby the government to impose barriers to trade.[90] There are two main kinds of barriers to trade we will consider: tariffs and quotas.

One very important result to keep in mind is that any price-quantity outcome that results from the imposition of a tariff could also be achieved by the imposition of *some* quota.[91] Similarly, any price-quantity outcome that results from the imposition of a quota could also be achieved by the imposition of *some* tariff.

14.4.1. Tariffs

A tariff is a per-unit tax imposed on imports. When a country imposes a tariff on a good, the price of that good rises in the importing country – which then of course means that the domestic quantity supplied and the domestic quantity demanded change from what they were in the absence of the tariff.[92]

Because the world market must still clear, the amount imported in the importing country must still equal the amount exported in the exporting country. In Figure 45 below, this means that line segment RS must be the same length as line segment QT:

[90] Indeed, this description seems to fit the events that led to President Bush imposing barriers to free trade in steel in March 2002, barriers to which other countries are beginning to retaliate.
[91] Not just *any* quota, however.
[92] The quantity demanded is lower and the quantity supplied is higher – which of course is simply because the demand curve is downward-sloping and the supply curve is upward-sloping.

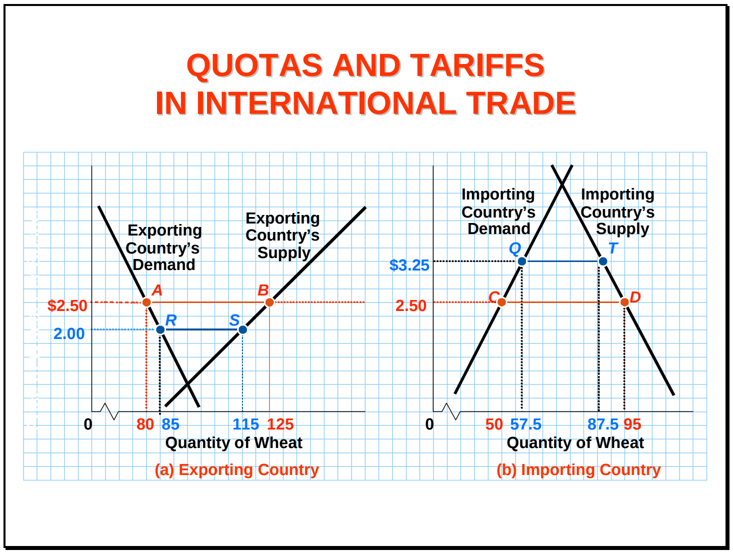

Figure 45

This then determines the price of the good in the exporting country. The price of the good in the exporting country falls below the free-trade world price, but is still higher than in autarky. We have thus completely determined the effects of the tariff in each country.

14.4.2. Quotas

A quota is a limit on the number of goods that can be imported. In the importing country, then, the supply curve simply shifts to the right by the amount of the quota, and the price in the importing country is then determined by the intersection of this new supply curve and the country's demand curve. The reason why the supply curve shifts out by the amount of the quota is that at every price, the total supply available in the importing country is the sum of the domestic supply and the (quota-limited) imports.

If the quota is set in a particular way, then the country can achieve the same outcome that occurred in Figure 45 under the tariff. The quota must be the amount such that the domestic supply curve shifts so that the intersection of the shifted supply curve and the country's demand curve occurs exactly at the price that includes the tariff in the importing country ($3.25 in Figure 45). Figure 46 illustrates this:

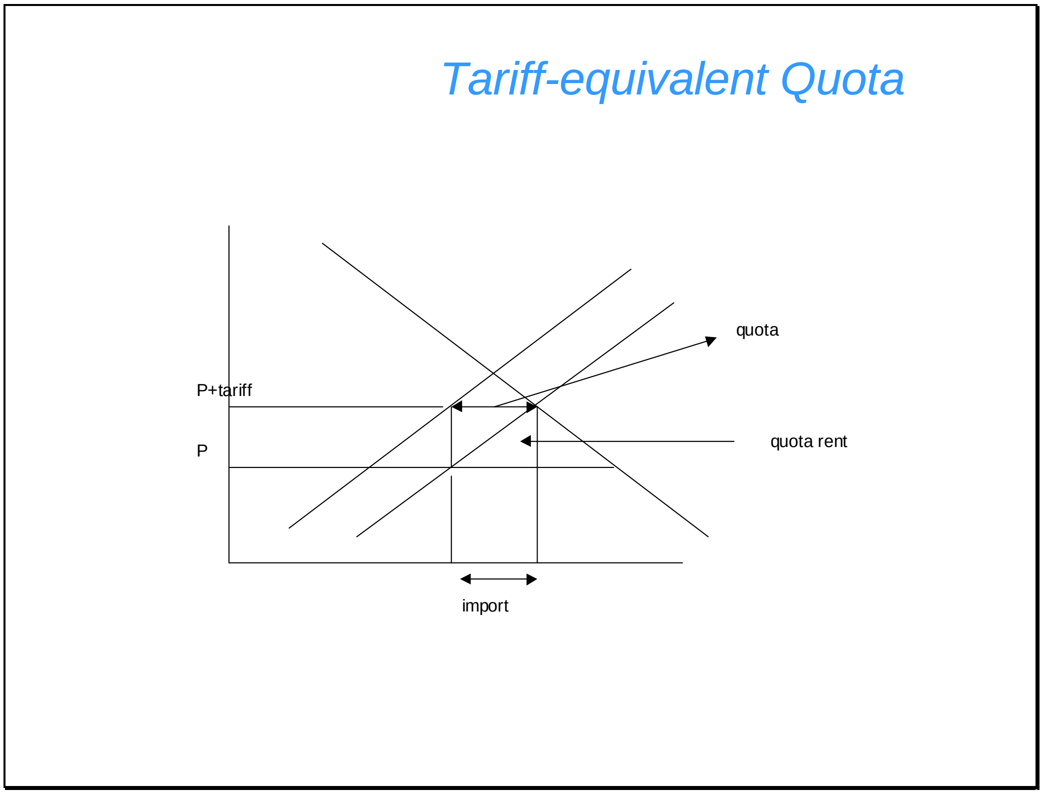

Figure 46

14.4.3. Rents

Note the rectangle marked "quota rent" in Figure 46 – this is the amount of extra revenue that importers receive in the importing country as a result of the quota. The intuition for why importers receive this rent is that the price per unit is higher under the trade restriction than if the quota were not in place.

Finally, notice that such a rectangle also exists in the importing country in Figure 45. The rectangle is the one that has top corners Q and T and bottom edge on the line segment CD. In that case (tariff), the rent goes to the government in the form of tariff revenue.

15. International Economics II: International Finance

The other major branch of international economics is international monetary economics, also known as international finance. Issues in international finance have importance for international trade as well as for the effectiveness and transmission mechanisms of both monetary policy and fiscal policy. Exchange rates between currencies play a central role in the field of international finance, so we will begin with a definition of exchange rates and then turn to a study of how exchange rates are determined, followed by a discussion of the two major types of exchange rate systems that exist between countries.

Very simply stated, the **nominal exchange rate** between two countries is the price of one country's currency in terms of the other country's currency. Because the nominal exchange rate is ultimately just a price, the main analytical tools to use in thinking about exchange rates are, as we will soon see, the familiar concepts of supply and demand.

Examples:

1. The exchange rate between the U.S. and Canada on January 4, 2002 was 1.59 Canadian dollars per one U.S. dollar. Another (completely equivalent!) way of stating this is that the exchange rate was 0.63 (=1/1.59) U.S. dollars per one Canadian dollar.
2. The exchange rate between the U.S. and Mexico on January 4, 2002 was 9.17 pesos per one U.S. dollar – or, equivalently, 0.109 (=1/9.17) U.S. dollars per one Mexican peso.
3. Immediately following the January 2002 devaluation of the Argentine peso, the exchange rate between the U.S. and Argentina was 1.4 pesos per one U.S. dollar – or, equivalently, 0.71 (=1/1.4) U.S. dollars per one Argentine peso.[93]

Very important to note is the two equivalent ways of expressing the exchange rate between two countries that is demonstrated in the above examples. Note that in each example, one way of expressing the exchange rate is simply the inverse of the other way of expressing it.

There are two major kinds of exchange rate regimes that exist in the world – floating regimes and fixed regimes.[94]

15.1. Floating Exchange Rates

If market forces are allowed to determine the exchange rate between two countries, the exchange rate is said to be "floating." As usual, "market forces" simply means the forces of supply and demand. You are already very familiar with what "demand for a good" and "supply of a good" means. Before proceeding, we should describe the factors that generate "demand for a currency" and "supply of a currency."

For the following discussion, assume there are two countries, country 1 and country 2, and the domestic currencies in these countries are currency 1 and currency 2, respectively.

15.1.1. Demand for a Currency

There are three reasons why the citizens of one country (country 1) would demand currency of another country (country 2). They are:

[93] By early April 2002, the exchange rate had moved further, to approximately 2.9 Argentine pesos per one U.S. dollar.
[94] The trend in recent history has been a move away from fixed exchange rates and toward floating rates.

1. Demand for goods and services produced in country 2
2. Demand for physical assets (i.e., factories) located in country 2
3. Demand for financial assets (i.e., stocks and bonds) of country 2

As would be expected, sellers (of goods, services, and assets) in country 2 want to be paid in their home currency, currency 2, no matter who is buying from them. Thus, when citizens of country 1 wish to purchase any goods, services, or assets from country 2, they will have to first buy currency 2 using their own currency, currency 1.[95] Thus, the above three channels all lead to demand for currency 2 by citizens of country 1. Because demand for a currency arises due to demand for other items, it is said to be a *derived demand.*

15.1.2. Supply of a Currency

Through the same channels as in the preceding discussion, citizens of country 2 will have a demand for currency 1. Thus, the citizens of country 2 will give up currency 2 in order to obtain currency 1. This surrender of currency 2 by citizens of country 2 leads to the supply of currency 2. That is, citizens of country 2 supply currency 2 to the exchange market because of their demand for the goods and assets of country 1. The demand for one currency, then, is intimately related to the supply of another currency. Indeed, we can think of the demand for a currency as essentially the same as the supply of another currency.

15.1.3. Equilibrium

The price of currency 2 in terms of currency 1, then, is determined as that price at which the demand for currency 2 equals the supply of currency 2,[96] illustrated in Figure 47.

[95] You may be thinking at this point that when you buy, say, a pair of shoes made in Italy at your local shop, you don't have any demand for lira (or, beginning January 1, 2002, euros). That is because some other agent has already made the necessary dollar for euro exchange for you – knowing that you (or somebody else) would eventually purchase. But it is perhaps easier and more instructive for you to think of the demander of euros in this case to be *you*, the consumer.

[96] So, clearly, the exchange market is just like any other market – price and quantity (in this case, exchange rate and amount of currency 2 bought and sold) are determined simultaneously.

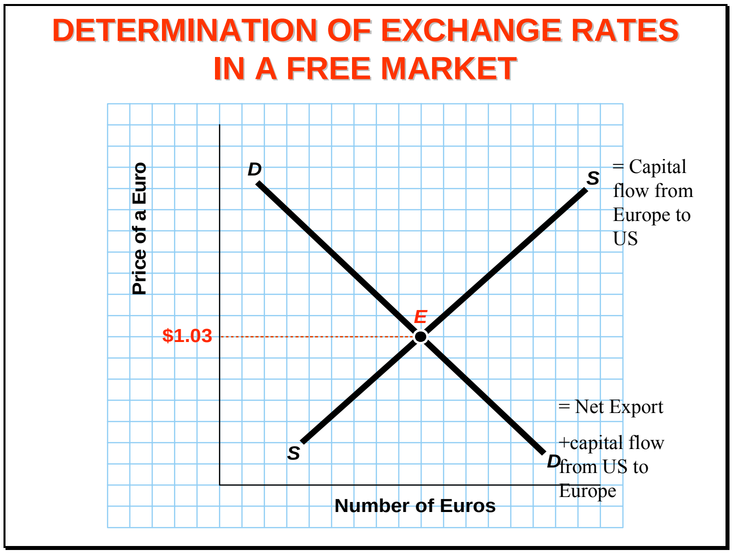

Figure 47

15.2. Currency Appreciation and Depreciation

Changes in floating exchange rates occur due changes in demand and supply. In currency markets, such price changes have a particular terminology associated with them. A currency is said to **appreciate** relative to another currency when it becomes more expensive in terms of another currency. In terms of the example in Figure 47, if the price of one euro rises, the euro has appreciated. A currency is said to **depreciate** when it becomes less expensive in terms of another currency. In terms of Figure 47, if the price of one euro falls, the euro has depreciated.

When considering any pair of currencies, if one currency appreciates versus the other, then *by definition* the second currency depreciates versus the first.

Example: Consider the following data:

EXCHANGE RATES WITH THE U.S. DOLLAR

			Cost in Dollars		
Country	Currency	Symbol	July 1980	Feb. 1985	June 1999
Australia	dollar	$	$1.16	$0.74	$0.66
Canada	dollar	$	0.87	0.74	0.68
France	franc	FF	0.25	0.10	0.16
Germany	mark	DM	0.57	0.30	0.53
Italy	lira	L	0.0012	0.00049	0.00053
Japan	yen	¥	0.0045	0.0038	0.0083
Mexico	new peso	$	44.0	5.0	0.10
Sweden	krona	Kr	0.24	0.11	0.12
Switzerland	franc	S.Fr.	0.62	0.36	0.65
United Kingdom	pound	£	2.37	1.10	1.60
--	euro	€	--	--	1.03

Figure 48

The table in Figure 48 gives the dollar price of several major currencies at three different points in time. Consider the Canadian dollar – between July 1980 and June 1999, the Canadian dollar depreciated versus the U.S. dollar. This then necessarily means that the U.S. dollar appreciated versus the Canadian dollar between July 1980 and June 1999.[97]

15.3. Factors Affecting Exchange Rates

Aside from the supply of and demand for currencies, the factors that are believed to affect exchange rates in general differ according to whether we are considering the long-run, the medium-run, or the short-run.

15.3.1. Purchasing Power Parity (PPP) in the Long-Run

In the long run, economists believe that the price in any country of a particular good should be the same, *once the prices have been put into the same currency units using the exchange rate*. This idea is known as purchasing power parity (PPP) and is very often used as a building block of research models of international finance.[98]

As an example of PPP, say the price level in country 1 (returning to the countries in our example above) is 500 (measured in currency 1) and the price level in country 2 is 1,000 (measured in currency 2). PPP then says that the exchange rate between the two currencies is 2 units of currency 2 for 1 unit of currency 1, which means that the price

[97] Make sure you understand which currency is appreciating and which currency is depreciating. This point often confuses people.

[98] There is actually a technical difference between PPP and something called the Law of One Price (LOOP), which is the proper term for the description just given of PPP. But for our purposes they are the same thing.

levels in the two countries are identical when measured in common units (i.e., the same currency).

15.3.2. The Medium Run

In the medium-run, a country with a relatively high growth rate will experience a depreciation of its currency. The mechanism by which this happens is that because of the high growth rate, consumption is growing very fast – which would (usually) be accompanied by fast growth of imports as well. The surge in demand for imports would then increase the demand for foreign currency (see the discussion above), thus resulting in a depreciation of the domestic currency.[99]

15.3.3. The Short-Run

In the short run, exchange rate movements are primarily determined by interest rate differentials among countries. Financial investors the world over who are seeking the highest possible returns will try, as much as possible, to acquire the financial assets of countries in which the interest rate is highest.

The mechanism by which this works is illustrated by the following example: say the interest rate in the U.S. is high relative to the interest rates in other countries. Then foreigners will want to hold U.S. financial assets because they offer better returns than the financial assets of other countries. This then means that the demand for U.S. dollars will increase, which causes the dollar to appreciate versus foreign currencies.

15.4. Fixed Exchange Rates

Some individuals and governments believe that floating exchange rates (that is, exchange rates determined by market forces) may have an excessively high degree of volatility – that is, the degree of uncertainty in floating exchange rates is undesirable for the macroeconomy. Under this line of thinking, this uncertainty[100] may then lead to less international trade than would exist if there were not so much uncertainty regarding exchange rates. The preceding reasoning sometimes motivates a government to *fix* its exchange rate vis-à-vis some other currency. That is, the government does not allow the exchange rate to be market-determined. Rather, it may peg the exchange rate at some level it believes is desirable.

To cast fixed exchange rates in language familiar from introductory microeconomics, the government is essentially establishing a price floor or a price ceiling when it decides to fix an exchange rate. The price floor (price ceiling) is implemented in this case of foreign exchange markets by the government buying (selling) the appropriate amount (the magnitude of the difference between supply and demand at the desired fixed

[99] And, of course, an *appreciation* of the foreign currency versus the domestic currency

[100] As well as the transactions costs associated with exchanging one currency for another, which was the primary motivation for the creation of the euro.

exchange rate) of its domestic currency. The means of payment (payment received) for the government comes out of (goes into) its official foreign reserves.[101]

15.4.1. Balance of Payments (BOP)

A few definitions are in order before proceeding to a more detailed consideration of how fixed exchange rate systems are managed.

Balance of Payments – a set of accounts that summarizes a country's transactions (real *and* financial) with the rest of the world; the main components of the balance of payments are the current account, the capital account, and the official settlements balance.
Current Account – measures the sum of net exports, net investment income, and other items.[102]
Capital Account – measures the net inflows of foreign monies; for the U.S., for example, it is the difference between "dollars that leave the country" and "dollars that enter the country" through international financial transactions.
Official Settlements Balance – records the change in official reserve assets (usually in the form of foreign currency or foreign bonds) held by the monetary authority of a country; if official reserves increase during some period of time, the official settlements balance for that period of time is *positive.*

For any country, the current account and capital account for any period of time sum to zero, a fact known as the **fundamental balance of payments identity.** You can think of the basic reason for this identity as arising from a country's need to finance a trade deficit (surplus) by borrowing (lending) abroad. Trade directly impacts the current account, and international financial transactions directly impact the capital account. Thus, if a single international transaction affects both the current and the capital accounts, it must be that it generates offsetting entries in the two accounts.[103]

The fundamental balance of payments identity involves the current account and the capital account. The official settlements balance is used to record changes in official reserve assets held by the monetary authority of a country. There are several types of transactions that are recorded in the official settlements balance, but the type of the most interest is interventions by the monetary authority in foreign exchange markets. If a government desires to maintain a fixed exchange rate of its currency vis-à-vis another currency, the mechanism by which it does so is buying or selling an appropriate amount

---

[101] Every government holds the currencies or bonds of (some or many) other countries as foreign reserves. These foreign reserves are used by the government for international financial operations, such as making interest payments to other countries and intervening in the foreign exchange market. U.S. bonds are the most widely-held foreign reserve asset.
[102] The main "other item" is an item called net unilateral transfers, which are gifts (of money or goods) from the U.S. to other nations. For example, when U.S. residents of Mexican origin send a total of $100 million to their relatives in Mexico, the current account receives an entry of negative $100 million.
[103] It is possible for an international financial transaction to impact only one of the accounts, in which case two offsetting entries are generated in the same account. However, the details of balance of payments accounting are beyond the scope of this text.

of its currency using its official reserves. The change in official reserves resulting from such intervention thus generates a nonzero balance of payments. The magnitude of the difference between the supply of a currency and the demand for that currency *is* what is termed the balance of payments (referred to from here on as BOP).[104]

When a country has a floating exchange rate regime, the BOP is necessarily zero – that is, the capital account and the current account balance each other,[105] and the official settlements balance is zero because the monetary authority's official foreign reserves were not used in foreign exchange interventions. However, under a fixed exchange rate regime, this is usually not the case. If the supply of a currency exceeds the demand for it, then the government must step in and absorb the excess supply. The way it absorbs the excess supply is by spending some of its foreign reserves to buy the excess currency. The country is then said to be running a BOP deficit. Figure 49 (the case of the Thai baht before its 1997 devaluation) illustrates the case of a BOP deficit induced by a fixed exchange rate that is higher than the equilibrium exchange rate.

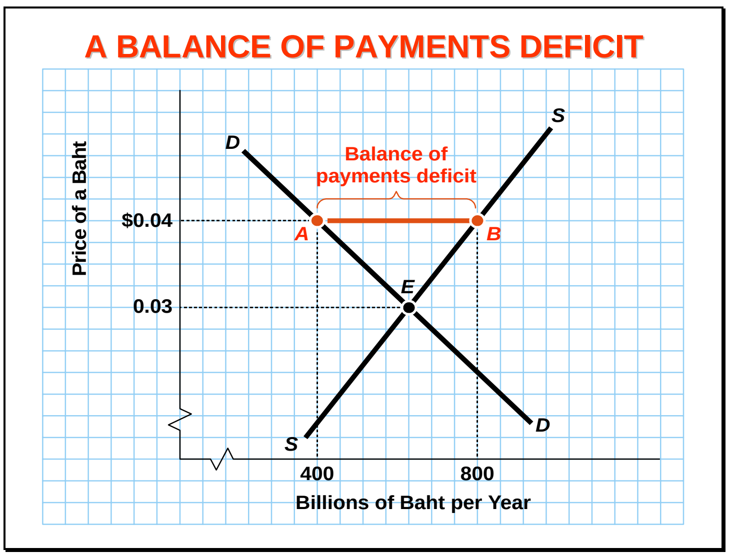

Figure 49

15.4.2. Speculative Runs

When a country is running a BOP deficit, it is draining its foreign reserves, which is the instrument by which the fixed exchange rate is maintained. If the BOP deficit persists, the country will eventually run out of foreign reserves. At that point, it will no longer be able to defend its fixed exchange rate, and the currency will become devalued. If this were the end of the story, then the exchange rate would fall back to its equilibrium level, point E in Figure 49.

[104] Note that the current account, capital account, the official settlements balance, and the BOP are all flow variables.
[105] At least in theory – there is usually a statistical discrepancy in the data, though, which usually is attributable to measurement errors.

But in currency markets, this is rarely the whole story. Because BOP figures and foreign reserve levels for a country are public information, people (financial investors) will know that a country is running out of foreign reserves before they are actually completely drained. In order to protect their own financial assets, rational investors will want to dump their currency holdings before the devaluation. Thus, there will be a large increase in the supply of the currency before its anticipated devaluation. The result will be that when the country actually does run out of reserves and has to let the currency fall to its equilibrium value, the equilibrium value will be *lower* than what it would have been if the currency had been allowed to float freely in the first place.[106] Figure 50 illustrates this phenomenon for the case of Thailand in 1997.

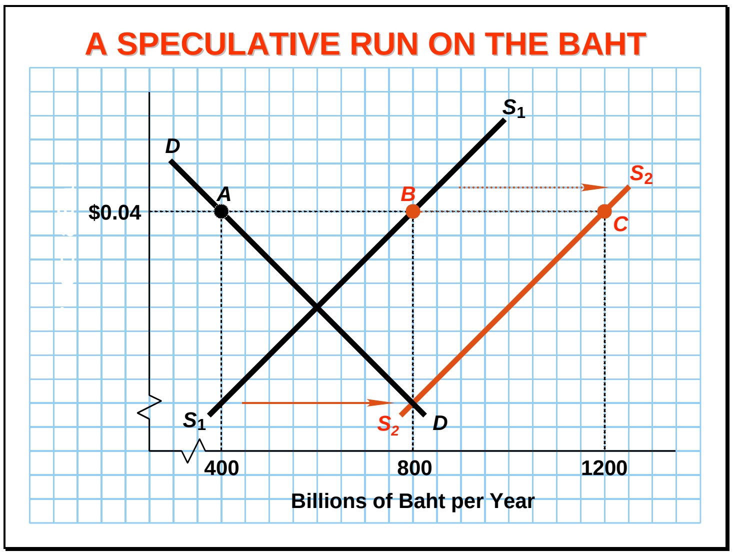

Figure 50

The surge in supply of the currency before foreign reserves run out is what is known as a speculative attack on the currency.

15.4.3. BOP Surplus

Although the above discussion has focused on BOP deficits (that is, on fixed exchange rates in which the peg is above the equilibrium exchange rate), a BOP surplus is a perfectly well-defined concept as well. A BOP surplus arises when a fixed exchange rate is at a level that is *below* the equilibrium exchange rate and is illustrated in Figure 51.

[106] Indeed, foreign reserves will end up being drained more quickly due to the speculative attack. Such an event is called a *balance-of-payments crisis,* the careful study of which is a topic for a more advanced course in international economics.

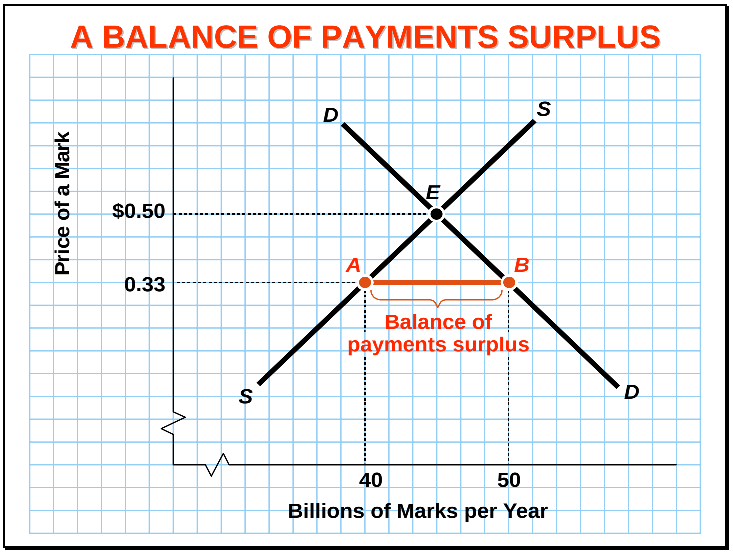
A BALANCE OF PAYMENTS SURPLUS
Price of a Mark
$0.50
0.33
D
S
E
A
B
Balance of
payments surplus
S
D
40
50
Billions of Marks per Year

Figure 51

16. International Economics III: Open Economy Macroeconomic Policy

We have now studied how economies are linked through international trade and international finance. We now reconsider the short-run channels through which fiscal policy and monetary policy affect an economy. The key variable in this reconsideration of macroeconomic policy is the nominal exchange rate. The reason that nominal exchange rates are crucial in this more detailed analysis of the effects of policy is that exchange rates affect relative prices of goods between countries. That is, currency appreciations and depreciations serve to change international relative prices, holding the price levels in the given countries fixed.[107] With this in mind, it is straightforward to analyze the impact of a currency appreciation or depreciation on aggregate demand and aggregate supply.

16.1. Macroeconomic Effects of Currency Appreciation

When a country's (call it country 1) currency appreciates (holding all else constant), foreigners find it more expensive to buy country 1's goods, services, and financial assets. Also, citizens of country 1 find it *cheaper* to buy the goods of other countries. For both of these reasons, net exports of country 1 decline, which shifts the AD curve in.

In addition, because in many developed countries many input goods for production are purchased from foreigners (the main example being oil), a currency appreciation makes the purchase of these input goods relatively *cheaper* for producers in country 1. Thus, for any price level, producers in country 1 can supply more output – which means the AS curve shifts outwards.

On balance, the inwards shift of the AD curve and the outward shift of the AS curve produces an ambiguous change in real GDP. Evidence shows, though, that the effect of a currency appreciation is more pronounced on AD than on AS – that is, the AD shifts in by more than the AS shifts out. Overall, then, the price level and real GDP in the domestic country both fall because of the appreciation of the domestic country, as illustrated in Figure 52.

16.2. Macroeconomic Effects of Currency Depreciation

The effects of a depreciation of the currency of country 1 are the reverse of those just described above. Following a depreciation of the domestic currency, the AD curve shifts outward because net exports unambiguously rise due to the depreciation. That is, the depreciation makes foreign goods more expensive for domestic residents, pushing down

[107] Note the implicit assumption here that purchasing power parity (PPP) does not hold. But recall that PPP is expected to hold in the long-run, not in the short-run. Macroeconomic stabilization policies are inherently short-run issues, so there is no "violation" of the PPP theory here.

imports, and domestic goods less expensive for foreigners, raising exports. The AS curve shifts inwards because foreign input goods are now more expensive for domestic producers. Thus, the overall effect of a nominal depreciation is a rise in the domestic price level and real GDP.

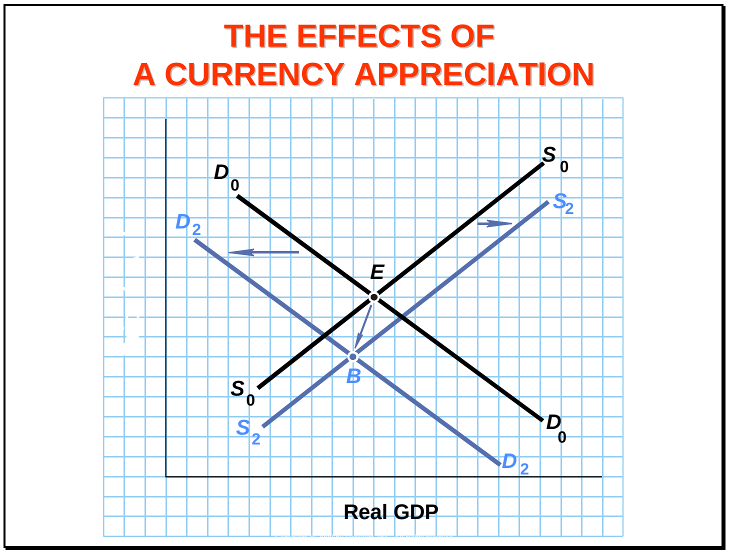

Figure 52

16.3. Impact of Interest Rates on Capital Flows

In our earlier study of the impact of interest rates on aggregate demand, we saw that a rise in the interest rate caused the AD curve to shift inwards, because higher interest rates lead to lower investment. Now we are ready to understand another mechanism by which interest rates impact AD, namely through the effect of interest rates on exchange rates.

Recall from earlier that the primary determinant of exchange rates in the short run is interest rates. That is, the higher the interest rate is in a country, the higher demand by foreigners is for the financial assets of that country. This in turn increases demand for that currency, which makes the currency of that country more valuable (that is, the currency appreciates).

But we just saw above that currency appreciation leads to a decline in aggregate demand because net exports fall. So this is the other channel through which a rise in interest rates leads to an inward shift of the AD curve once we take account of the international sector. That is, if a country were in autarky, an increase in interest rates would have *less* of an adverse effect on AD than if it were open to international trade.

We are now ready to re-analyze the transmission mechanisms of fiscal and monetary policies. We will do so by analyzing two specific examples – an expansionary fiscal policy and, separately, an expansionary monetary policy. You should be able to trace the effects of a contractionary fiscal policy and a contractionary monetary policy yourself.

16.4. Expansionary Fiscal Policy

Recall that a government conducts expansionary fiscal policy by increasing government spending, reducing taxes, or both. Thus, the AD curve shifts out as a result of expansionary fiscal policy, simply because G and T appear as components of the fundamental identity $Y = C + I + G + NX$.[108] As long as the AS curve stays fixed (that is, we are ignoring any potential supply-side effects here), both the price level P and real GDP increase.

There is also an impact on the domestic interest rate as a result of fiscal policy, which we noted briefly in our discussions of Ricardian Equivalence and the crowding-out effect of fiscal policy. Recall from our study of the domestic funds market that national savings is potentially affected by fiscal policy. If we discard the assumption of lump-sum taxes and suppose that all taxes are distortionary, as in reality, then we concluded that Ricardian Equivalence does not apply at all, and *any* fiscal policy must affect the national savings schedule. Thus, a fiscal expansion shifts the national savings schedule inwards, and, with the investment schedule fixed, the *closed-economy real interest rate* rises.

In an *open* economy, however, it might seem to you that what the closed economy interest rate *would* be should not matter at all. In reality, no economy is completely open nor completely closed. Thus, the actual interest rate that prevails in an economy *is* influenced to some degree by what the closed economy interest rate *would* be. Putting the above discussion together, *an expansionary fiscal policy raises the dometic interest rate.*[109]

There is an alternative way to see that fiscal policy affects the interest rate. Recall that the money demand function $M^D = L(PY, i)$ states that money demand depends positively on the volume of transactions (which is proxied by PY) and negatively on the interest rate. Because both P and Y increase due to expansionary fiscal policy, the money demand function tells us that the M^D curve shifts out in a graph of interest rate versus quantity of money.[110] As long as the money supply curve stays fixed, the interest rate rises.

Now we can understand the international linkage. Interest rates are the primary determinant of exchange rates in the short run. Thus a higher interest rate, because it makes domestic financial assets more attractive to foreign investors, induces an appreciation of the domestic currency. That is, if interest rates rise in country 1, the currency of country 1 appreciates.

[108] The T appears, of course, in the consumption function.
[109] With inflation expectations fixed in the short-run, it is both the real interest rate and the nominal interest rate which rise.
[110] This mechanism should be familiar from our earlier study of money demand.

The effect of a currency appreciation is that it raises the relative price of goods of country 1 relative to the goods of the rest of the world – or, stated equivalently, the relative price of foreign goods relative to the goods of country 1 decrease. But we have already seen what happens as a result of this event. Net exports of country 1 decrease, which shifts the AD curve inward. On the other hand, the AS curve shifts outward due to the fact that imported inputs are cheaper.

So there are two effects on the AD curve as a result of expansionary fiscal policy. The direct effect is that the AD curve shifts out due to the Keynesian multiplier. But the AD curve also shifts in because of the effect on interest rates and thus on exchange rates. On balance, the AD curve does shift out from its initial position – just not as much as it would if the economy were closed. The overall effect, then, of expansionary fiscal policy is as illustrated in Figure 53.

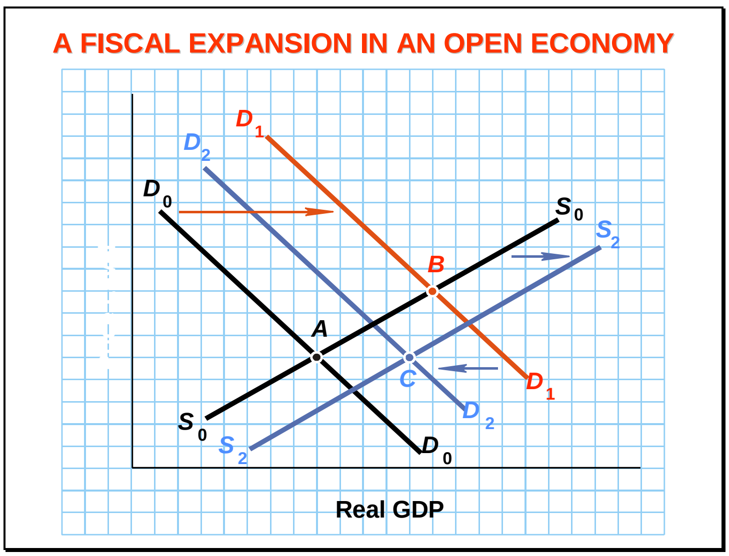

Figure 53

The most important point to take away from this discussion is: **the power of fiscal policy is weakened because of international capital flows.**

16.5. Expansionary Monetary Policy

Much more direct to understand is the effect of an expansionary monetary policy on interest rates – in fact, we have already studied this in depth. An expansionary monetary policy shifts money supply outward (assume that banks hold no excess reserves and there is no currency drain, for simplicity) – thus reducing interest rates.

Lower interest rates induce higher investment – which pushes the AD curve outward, since investment is a component of the fundamental accounting identity. We already know this mechanism from our earlier study.

But now, because of international capital flows, we know that lower interest rates in country 1 will induce a depreciation of the currency of country 1. This depreciation makes the goods in country 1 relatively less expensive for foreigners. Thus, net exports will *rise*, pushing the AD curve even further out. Also, the depreciation will cause the AS curve to shift in because imported inputs will now be relatively more expensive for producers in country 1, but recall from earlier that the effect on AS is smaller than the effect on AD.

We see here that *both* effects on the AD curve are that the AD curve shifts out. Thus, **the power of monetary policy is enhanced because of international capital flows.**

16.6. The "Twin Deficits"

Here we visit an issue using tools we have already developed – the relationship between the government budget deficit and the trade deficit.

In the 1980's the U.S. ran large government budget deficits and large trade deficits simultaneously. These came to be known as the "twin deficits." These twin deficits, however, are actually not unrelated to each other, as we shall now see. We start with some algebraic manipulations:

$$S^{priv} = Y - T - C$$

where, recall, S^{priv} is private savings. Using the accounting identity $Y = C + I + G + NX$ to substitute for Y in the preceding expression, we get

$$S^{priv} = I + G + NX - T$$

which, after further manipulation, yields

$$(S^{priv} - I) + (T - G) = NX$$

Finally, recalling that government savings is $S^{gov} = T - G$, we can write the last expression as

$$S^{priv} - I + S^{gov} = NX$$

This expression is simply the funds market equilibrium condition that we have already studied, except now we allow net exports to be non-zero. The above expression states that the trade balance is equal to the sum of the government budget balance (savings of the government) and the difference between private savings and investment.

Looking at data for savings and investment for the U.S. in the 1980's, we find that, roughly, $S^{priv} = I$. Thus, as an approximation to what occurred in the U.S. in the 1980's, we can write

$$S^{gov} = NX .$$

But this expression simply states that if the government is running a budget deficit, then the nation *must* be running a trade deficit – and vice versa. Thus, the two "twin" deficits of the 1980's were really the same phenomenon.

The situation in the U.S. in the 1990's was somewhat different. The U.S. continued to run large trade deficits in the 1990's – however, the government budget came into balance.[111] The explanation of this comes from the savings and investment channels. Whereas in the 1980's $S^{priv} = I$ was roughly true, in the 1990's private savings declined, whereas investment soared. Thus the term $S^{priv} - I$ was negative for much of the 1990's – which explains why the trade balance (NX) was negative. The lesson here is that the trade position of a country and the government budget position are related to each other, but they do not *necessarily* have the same root cause.

The fact that the U.S. was borrowing to fund its rapidly growing investment in the 1990's is a point that many commentators have noted as an unsustainable driving force of the U.S. economic boom of the past decade. At some point, the trade deficit will have to correct itself because foreigners will eventually decide to stop sending their financial capital to the U.S. in such massive amounts. However, it is unlikely that this will happen in the near future because the U.S. remains the strongest/safest place to invest in the world, and as such, correcting the trade deficit should probably not be given such high priority as some observers would like to give it.

Nevertheless, the main ways to cure a trade deficit are:

1. Tighter fiscal policy and looser monetary policy – both promote net exports through their downward pressure on interest rates.
2. Increased economic growth abroad – the better other countries are doing, the more they will want to import from other countries, thus net exports in those other countries rise.
3. Increased national savings and/or decreased investment.
4. Protectionism – however it is very possible that other nations would impose barriers to trade in retaliation, and economic theory tells us this is not a very good solution.

[111] Actually, as you probably know, it went into surplus for a time – but let's just say for argument's sake it only became exactly balanced. Of course, more recently, the fiscal budget has fallen back into deficit.

16.7. The Asian Financial Crisis

For more than two decades, the economies of East Asia all grew very rapidly – these economies came to be called the "Asian tigers." Until the early 1990's, this rapid growth was financed mostly by local savings. However, in the 1990's foreign investors began investing much more heavily in the East Asian economies.

Because many Asian economies were pegging their currencies to the U.S. dollar (in particular, Thailand) at a rate higher than the equilibrium exchange rate, their persistent balance of payments deficits were eroding their foreign reserves. Eventually, investors realized that the Asian governments were not going to be able to defend their currencies much longer – which led to speculative attacks. Thailand's currency, the baht, was the first of the East Asian currencies to be attacked – as a consequence, the government of Thailand was forced to devalue its currency in July 1997, and the value of the baht depreciated approximately 50% quite rapidly.

One effect we have learned should happen in the face of currency depreciation is that net exports should be given a boost. However, there is another effect we have not mentioned yet, which profoundly affected the outcome of the Asian crises.

Much of the debt held in these economies was dollar-denominated – meaning that loans had to be paid back in dollars. Because of the large currency depreciations, however, dollars became much more expensive. Thus, the baht- (or other local currency-) values of debt increased significantly, causing much hardship for debtors in these economies.

There are many other factors to consider in describing the causes and effects of the Asian currency crisis, mostly focusing on inadequate preparation to deal with the transition to market-based capitalism. The above discussion provides only a brief introduction to the subject.

17. Economic Growth

We now return to an issue that we hinted at in the very beginning of our consideration of macroeconomics. Whereas much of our studies have focused on short-run macroeconomic fluctuations, the topic of economic growth is concerned with how an economy develops over long periods of time. Although it can be criticized on several fronts, the main measure of economic standard of living used is real GDP per capita (i.e., per person). Thus, economic growth is measured as growth in real GDP per capita. For much of history, there was essentially zero economic growth. But over the past two to three centuries, economic growth has been positive, although growth rates still vary widely from one region of the world to another and from one country to another.

There has been much debate about whether growth rates should imply "convergence" of real GDP per capita across countries. The notion of convergence is simply that over long periods of time, per capita GDP should become equal in even widely-differing countries. Evidence on this issue has been mixed. Many industrialized nations have indeed more or less converged to the same standard of living. However, many developing countries seem to be "stuck" at far lower standards of living, though some evidence of "catch-up" has been witnessed in Asian economies.

We have already studied growth accounting in detail – so we will immediately proceed here to the study of one of the most important theoretical models of economic growth.

17.1. Neoclassical Growth Model (aka Solow Growth Model)

An important theoretical model of economic growth is the neoclassical growth model, also known as the Solow model. In modern macroeconomics, the neoclassical growth model actually underlies the study of short-run as well as long-run phenomenon.[112] As will become clear, this model is really a model of how and why an economy accumulates physical capital. There are a number of assumptions we will make in studying the dynamics and implications of this model, assumptions that can be rightly criticized as being too unrealistic. However, the model is still very valuable to study in that it can provide a useful benchmark for the performance of other models of growth.

The assumptions we will employ in our study of the neoclassical growth model are:

1. Ignore business cycle fluctuations – the main implication of this is that hours worked in the economy are constant (that is, people don't derive utility from leisure, thus there is no decision made about how many hours to work).
2. No population growth (this will not affect any of the results of the model).

[112] Although we have not considered how the neoclassical growth model can be used to study short-run macroeconomic fluctuations, much of modern macroeconomic research of short-run phenomenon uses it as a starting point. The topic of how this is done is a topic for a more advanced course in macroeconomics.

3. National Savings = Investment (thus we ignore any issues about the government budget and the trade balance).
4. Individuals save a constant fraction of output.
5. Capital depreciates at a constant rate.

As before, we will continue to use the Cobb-Douglas production function, $Y = AK^{1/3}N^{2/3}$. However, with the assumption of zero population growth, we can rewrite the production function using per capita (per person) variables. Lowercase letters will denote per capita variables – in particular,

$$y \equiv \frac{Y}{N}, \; k \equiv \frac{K}{N}$$

Using these two definitions, the production function can be manipulated as follows:

$$Y = AK^{1/3}N^{2/3}$$
$$\frac{Y}{N} = \frac{AK^{1/3}N^{2/3}}{N}$$
$$y = A\left(\frac{K}{N}\right)^{1/3}$$
$$y = Ak^{1/3}$$

This last expression, $y = Ak^{1/3}$, is the version of the production function we will use – it expresses per capita output as a function of the per capita capital stock. Note that the function displays diminishing marginal product in capital. That is, the shape of the production function is as in Figure 54.

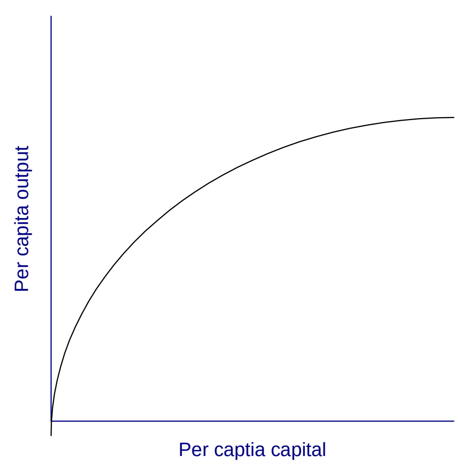

Figure 54

A few remarks about assumptions 4 and 5 above are in order. If individuals save a constant fraction of output, then in the context of the per capita production function we are using, this implies that savings = $sy = sAk^{1/3}$, where s is the constant savings rate between zero and one. This means that the savings function is simply a scaled version of the production function – in particular, scaled by the factor s. This is seen below in Figure 55.

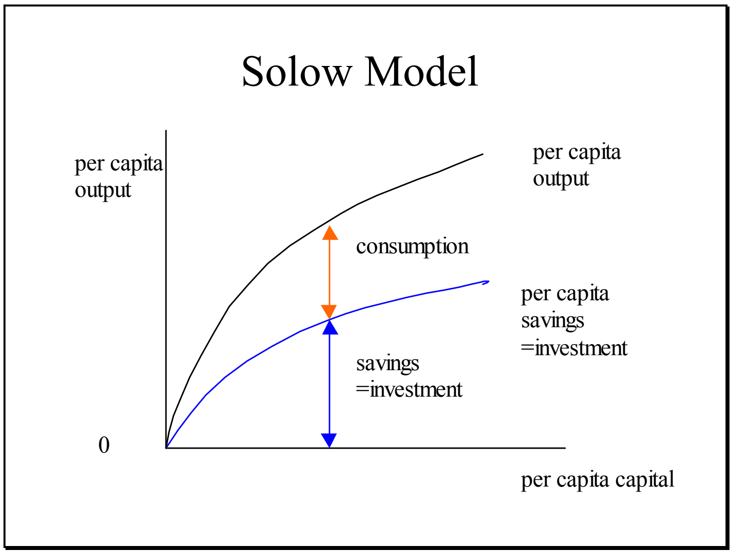

Figure 55

The assumption of capital depreciation is a natural one. Physical equipment naturally wears down over time as it is used in production processes. This is the notion that is captured in assumption 5 above. The assumption of constant depreciation is a decent approximation to what empirical studies suggest is the annual rate of depreciation of the U.S. capital stock, approximately 8% per year. Mathematically, then, depreciation is expressed as

$$depreciation = \delta k$$

where δ (the Greek letter "delta") is the constant rate of depreciation. So total depreciation is simply a linear function of the per capita capital stock.

Figure 56 displays both the savings function and the depreciation function. You should think of the depreciation function as that amount of investment required to simply maintain the capital stock from one year to the next (i.e., zero net change in the capital stock) – for this reason, we will sometimes call it replacement investment.

The intersection of the depreciation function and the savings function defines the steady state capital stock, denoted k*. The steady state capital stock is that level of the capital stock at which the amount of actual savings in the economy (which, recall by assumption 3 above, equals the amount of investment) equals the amount of replacement investment. That is, when the capital stock is k*, there will be no further change in k.

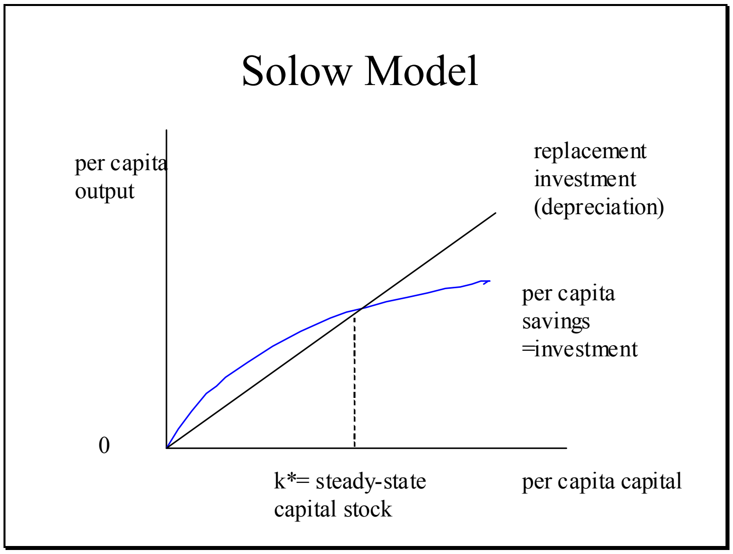

Figure 56

17.1.1. Dynamics of Growth

Now we are ready to analyze the dynamics of growth in the neoclassical growth model, and the workhorse of the analysis will be Figure 56.

Suppose that in the current period (call it period t) the economy has a per capita capital stock that is below k*. In this case, actual investment is larger than replacement investment, which means that in the next period (call it period t +1), the capital stock will be larger than in the current period. This is seen in Figure 57.

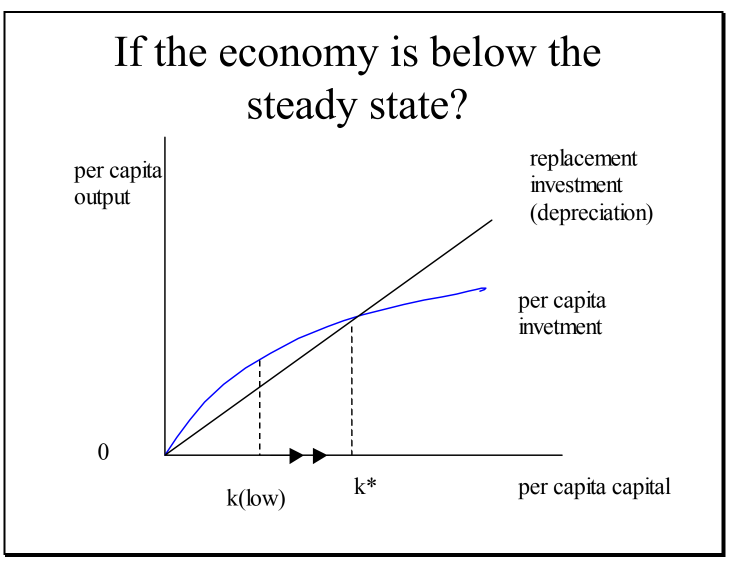

Figure 57

In period t + 1, the capital stock will still be below k*, but it will be closer to k* than it was in period t. At this new capital stock in period t +1, actual investment will still be larger than replacement investment (just look again at Figure 57) – but now the magnitude of the difference between actual and replacement investment[113] is not as large. Once again, the capital stock rises, and in period t +2, the capital stock is closer yet to k*. This then implies that the capital stock will once again rise in period t + 3. But this rise will be even smaller than the rise which occurred in period t + 2 – the simple reason for this is that the difference between actual investment and replacement investment is smaller yet. This argument repeats itself until the capital stock is equal to its steady state value.[114] The mechanism by which the capital stock *decreases* until it hits steady state if it starts *above* k* is analogous – the above logic simply operates in reverse.[115]

17.1.2. Convergence Across Countries?

By the discussion above, there is an inverse theoretical relationship between the level of the capital stock and the growth rate of the economy – the lower is the capital stock, the faster the economy should grow. Equivalently, because the production function $Ak^{1/3}$ is a monotonic function of per capita capital, we can state this correlation as a lower per capita GDP implies a higher growth rate of per capita GDP. This is one main prediction of the neoclassical growth model. However, empirical evidence does not bear this conclusion out, as Figure 58 shows. Figure 58 plots per capita GDP in 1950 for a number of countries (both developed and developing) versus the realized growth rate of per capita GDP in those countries over the subsequent 31 years. The prediction of the neoclassical growth model is that we should see an inverse relationship – basically, all countries should line up along the solid downward-sloping curve in Figure 58. As we can see, this turns out not to be the case.

[113] The difference between actual investment and replacement investment is termed net investment – it is net investment that leads to an increase in the capital stock over time.

[114] Actually, to be completely precise, the capital stock theoretically never actually reaches steady state – because in every period only part of the remaining difference from the steady state is made up. But for practical purposes, it is easy to be convinced that the steady state actually will be achieved.

[115] You should work though this argument yourself.

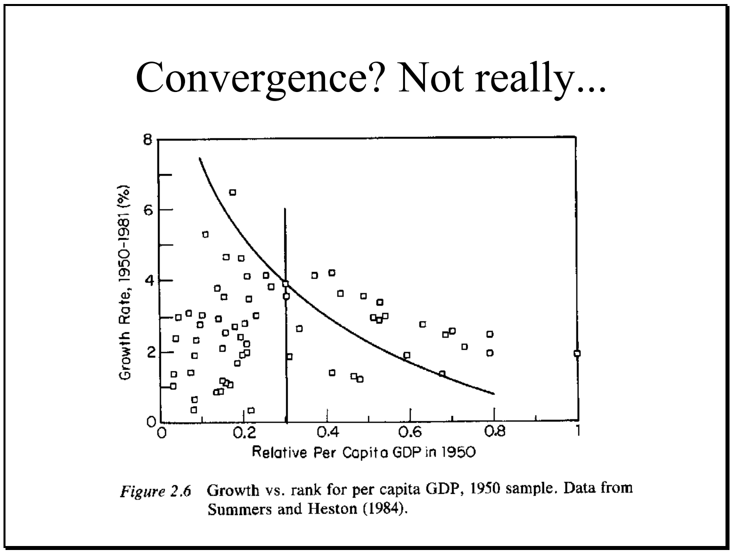

Figure 2.6 Growth vs. rank for per capita GDP, 1950 sample. Data from Summers and Heston (1984).

Figure 58

17.1.3. Shortcomings of the Neoclassical Growth Model

One major shortcoming of the neoclassical model is that it predicts that all countries should eventually converge to the same steady state. If one believes that "mature" economies (such as the U.S. and Europe) have all converged to their steady states,[116] then we should observe identical per capita capital stocks in all mature economies. But this is not what evidence shows.

One plausible modification to the theoretical model that allows this is to allow different countries to have different savings rates. That is, even if economies have the same production technology, perhaps one country has savings = $s_1 y$ and another has savings = $s_2 y$. As long as they both have the same depreciation rates, the steady state level of capital in the two economies will be different, as seen in Figure 59.

[116] Which at first blush is not a terrible thought – these economies have had a lot of time to converge to their supposed steady states.

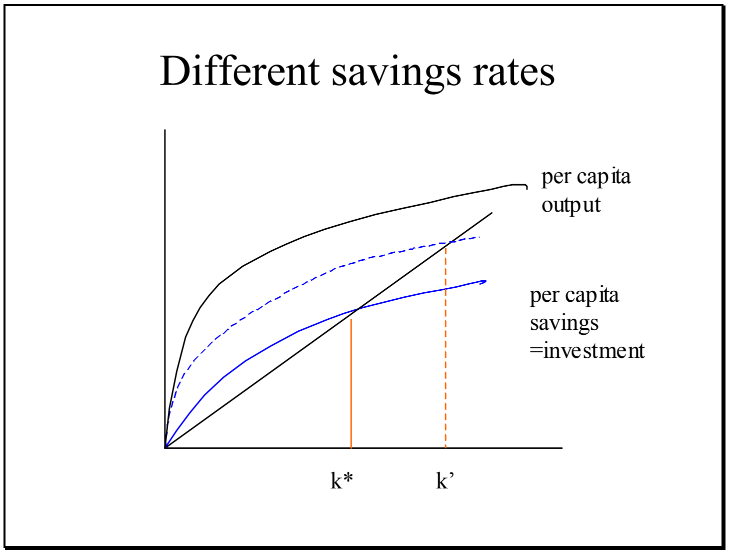

Figure 59

Another feature of the model that could be modified is to allow for different depreciation rates of capital in different economies. You should consider for yourself how such a modification to the model would work.

Another major shortcoming of the model is that it predicts that economies eventually reach a state of zero growth. That is, once steady state is achieved, there is no further economic growth because actual investment always equals replacement investment from that point on. In mature economies, such as the U.S., it is not hard to believe that the economy *should* have reached steady state by now because the economy has been in operation for so long. Or at the very least, growth rates should be clearly slowing down as the economy approaches steady state.

Empirically, it is certainly not true that the U.S. has permanently stopped growing – the recent downturn of the U.S economy notwithstanding. Nor is it empirically supported that growth rates are declining. If these previous two facts are true, then the neoclassical model of growth as we have considered it cannot account for all the facts.

The major modification to the model needed to correct this shortcoming is to allow for increases in the technology parameter A in the production function. Increases in A would shift the production function up (in a non-parallel manner, of course – since zero input would still have to yield zero output). A shift up in the production function would then mean a shift up in the savings function, *even if the savings RATE s is held constant.* This can be seen in Figure 60.

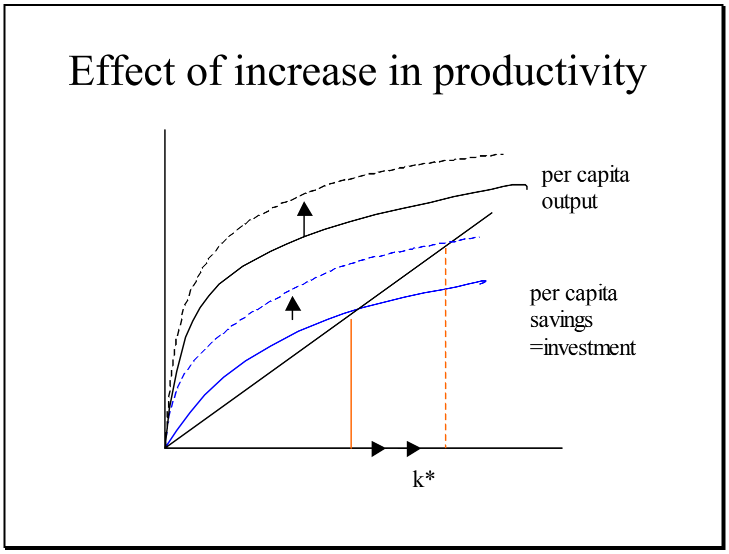

Figure 60

The neoclassical growth model can be "fixed" in this way, by imposing technical progress to explain continued growth in mature economies. However, the model does not have anything to say about *why* this advance in technology occurs – and any satisfying model of growth should be able to say something about where the driving force of the growth comes from. This is the subject of endogenous growth theory, to which we will turn next.

17.2. Endogenous Growth Theory (aka New Growth Theory) (OPTIONAL)

The main focus of endogenous growth theory is to offer explanations to address the shortcomings of the Solow model, namely that in the long run economic growth ceases – or, if it does continue to occur, it occurs because of some unexplained change in the state of technology.

The main concept that endogenous growth theory applies to "fix up" the Solow model is that there exist positive externalities[117] to innovation. For example, when a firm develops a new method of writing software, it will benefit that firm directly because of increased sales and the customers of the firm will benefit because of the new product. However, other firms in the economy, simply by being exposed to the new ideas generated by the innovating firm, will benefit as well. The exposure to new ideas will presumably enhance their design and manufacture, etc. of new products – which in turn will help yet other consumers and lead to more ideas available for yet other firms to use. Note that these effects will occur even if innovating firms are granted patents or copyrights for their inventions and development. A patent or copyright indeed grants certain rights to its holder – but it cannot prevent the dissemination of *ideas* through an economy, and it is ideas that fundamentally drive technological progress. Thus, in the language of the

[117] You should be familiar with the notion of externalities from basic microeconomics.

Solow growth model, the technology parameter A increases over time – implying economic growth even in the long run.

However, a firm, when deciding how much input to use in the development of new products or services, will not take into account the positive externalities of its innovations. In the language of the theory of externalities, the private (i.e., to the firm) marginal benefit from innovation is smaller than the social marginal benefit. Or, another way of stating this is that the private marginal cost of innovation is larger than the social marginal cost of innovation. Thus, the amount of resources that a firm will use for research and development purposes will be smaller than the amount of resources that it *should* use *if* it cared about maximizing the welfare of the entire economy.[118]

The above discussion leads to a very important point: there is clearly a role for government intervention in promoting innovation. Assuming that governments *do* care about maximizing the welfare of its citizens (even when private firms seek only to maximize their own profits), various policies can be implemented which encourage the socially optimal amount of innovation to occur. The most obvious in the context of the above example is for government to use public funds to subsidize research and development. Such a policy would have the effect of lowering the private marginal cost of research – which then induces the firm to engage in the socially optimal amount of research and development.

Because ideas can disseminate through the economy in the manner described, the original innovating firm cannot rest on its laurels. It will know that other firms will soon try to copy its products and enhance them – which will spur the original firm to continue developing new ideas. Thus, in this manner, the state of knowledge continually evolves.

Other ways that governments can encourage technical progress are through encouraging international trade and improving the quality and quantity of education. Again, both of these policies would expose domestic economic agents to more ideas – the ultimate engine of economic growth.

[118] Note that this does not necessarily imply that corporations are “evil” – they simply act to maximize their own private gain.