

IB Economics SL & HL

The complete set of class notes for the IB Economics student

JASON WELKER

THE ECONOMICS CLASSROOM

EconClassroom.com

Copyright © 2018 Jason Welker

All rights reserved.

ISBN: 1482612178
ISBN-13: 978-1482612172

DEDICATION

For Libby

TABLE OF CONTENTS

<u>Introduction to the Series</u>	7	The meaning of externalities	82
	8	Negative externalities of production and consumption	83
<u>1.0 Introduction to Economics</u>	8	Positive externalities of production and consumption	91
Economics as a social science	11	Lack of public goods	93
Scarcity	13	Common access resources and the threat to sustainability	94
Choice and Opportunity Cost	18	Asymmetric information (HL only)	95
Central Themes	20	Abuse of monopoly power (HL only)	96
<u>1.1 Competitive Markets, Demand and Supply</u>	20		
Markets	21	<u>1.5.1 Production and Costs (HL only)</u>	100
The law of demand	22	Production in the short run: the law of diminishing returns	101
The demand curve	24	Costs of production in the short run	104
The non-price determinants of demand and movements along vs. shifts of the demand curve	27	Costs of production in the long run	109
Linear demand equations, demand schedules and graphs (HL only)	31	Total revenue, average revenue and marginal revenue	111
The law of supply	32	Economic profit and normal profit	114
The supply curve	33	Profit maximization	116
The non-price determinants of supply and movements along vs. shifts of the supply curve	36	Alternative goals of firms	116
Linear supply equations and graphs (HL only)	40		
Equilibrium and changes to equilibrium	46	<u>1.5.2 Perfect Competition</u>	120
Calculating and illustrating equilibrium using linear equations (HL only)	50	Assumptions of the Perfectly Competitive Market Model	120
Consumer surplus	51	Revenue curves for the perfect competitor	121
Producer surplus	52	Profit maximization in the short run	122
Allocative Efficiency	56	Profit maximization in the long run	126
	56	Shutdown price and break-even price	130
<u>1.2 Elasticities</u>	56	Efficiency in perfectly competitive markets	131
Price elasticity of demand and its determinants	58		
Applications of price elasticity of demand	63	<u>1.5.3 Monopoly</u>	134
Cross price elasticity of demand and its determinants	64	Assumptions of the monopoly model	135
Income elasticity of demand and its determinants	65	Barriers to entry	136
Price elasticity of supply and its determinants	68	Revenue curves for the monopolist	136
	68	Profit maximization in monopoly	139
<u>1.3 Government Intervention in Markets</u>	71	Revenue maximization	141
Specific and ad valorem taxes	73	Natural monopoly	142
Tax incidence and price elasticity of demand and supply (HL only)	76	Monopoly and efficiency	147
Subsidies	77		
Price ceilings	81	<u>1.5.4 Monopolistic Competition and Oligopoly</u>	151
Price floors	81	Assumptions of the monopolistic competition model	151
<u>1.4 Market Failure</u>	81	Revenue curves for the monopolistic competitor	152
The meaning of market failure	81	Profit maximization in the short run	152
	81	Profit maximization in the long run	153

Non-price competition	154	Indicators of income equality/inequality	253
Monopolistic competition and efficiency	156	Poverty	257
Monopolistic competition compared with perfect competition and monopoly	156	The role of taxation in promoting equity	257
Assumptions of the oligopoly market model	158	Other measures to promote equity	261
Game theory	160	The relationship between equity and efficiency	262
Open/formal collusion	161	2.4 Fiscal Policy	263
Tacit/informal collusion	162	The government budget - sources of government revenue	263
Non-collusive oligopolies	162	Types of government expenditures	264
Price discrimination	169	The budget outcome	264
2.1 The Overall Level of Economic Activity	175	Fiscal policy and short-term demand management	266
The circular flow of income model	175	The impact of automatic stabilizers	272
Measures of economic activity: gross domestic product (GDP), and gross national product (GNP) or gross national income (GNI)	177	Evaluation of fiscal policy	274
The Business Cycle	184	2.5 Monetary Policy	284
2.2 Aggregate Demand and Aggregate Supply	188	Interest rate determination and the role of a central bank	284
The aggregate demand curve	188	Monetary policy and short-term demand management	288
The components of aggregate demand	190	Monetary policy and inflation targeting	294
The determinants of AD or causes of shifts in the AD curve	191	Evaluation of monetary policy	295
The meaning of aggregate supply	195	2.6 Supply-side Policies	297
Alternative views of aggregate supply	199	Supply-side policies and the economy	297
Shifting the aggregate supply curve over the long term	202	Interventionist supply-side policies - Investment in human capital	298
Short-run equilibrium in the AD/AS model	204	Interventionist supply-side policies - investment in new technology	298
Long-run equilibrium in the AD/AS model	209	Interventionist supply-side policies - Investment in infrastructure	298
The Keynesian spending multiplier (HL only)	215	Interventionist supply-side policies - Industrial policies	299
2.3 The Macroeconomic Objectives	221	Market-based supply-side policies - increased competition	299
Objective #1 - Low unemployment	221	Market-based supply-side policies - Labor market reforms	300
Consequences of unemployment	223	Evaluation of supply-side policies	301
Types and causes of unemployment	224	3.1 International Trade	303
Objective #2 - Low inflation: The meaning of inflation, disinflation and deflation	229	The benefits of international trade	303
Shortcomings of the inflation rate as a macroeconomic measure	232	Restrictions on free trade - types of trade protection	309
Consequences of inflation	235	Arguments for and against trade protection	315
Consequences of deflation	236	3.2 Exchange Rates	318
Types and causes of inflation	237	Freely floating exchange rates	318
Possible relationships between unemployment and inflation (HL only)	239	HL only objectives	323
Objective #3 - Economic growth	246	Causes of changes in the exchange rate	326
Consequences of economic growth	251		
Objective #4 - Equity in distribution of income	252		

Government intervention in forex markets - fixed exchange rates	331	4.5 The Role of Foreign Direct Investment (FDI)	408
Managed exchange rate systems	334	The meaning of FDI and multinational corporations (MNCs)	408
Evaluation of different exchange rate systems	337	Advantages and disadvantages of FDI for economically less developed countries	410
3.3 The balance of payments	339		
The meaning of the balance of payments	339		
The components of the balance of payments accounts	341	4.6 The Roles of Foreign Aid and Multilateral Development Assistance	412
The relationship between the accounts	343	Classifications and types of aid	412
The relationship between the current account and the exchange rate	344	Evaluation of foreign aid	418
Implications of persistent current account deficit (HL only)	346	4.7 The Role of International Debt	421
Methods to correct a persistent current account deficit (HL only)	347	Foreign debt and its consequences	421
The Marshall-Lerner condition and the J-curve effect (HL only)	348	4.8 The Balance Between Markets and Intervention	426
Implications of a persistent current account surplus (HL only)	350	Strengths of market-oriented policies	426
3.4 Economic Integration	351	Weaknesses of market-oriented policies	427
Preferential trade agreements	351	Strengths of interventionist policies	428
Trading blocs	353	Weaknesses of interventionist policies	429
Monetary union	355	Market with government intervention	429
3.5 Terms of Trade (HL only)	359		
Meaning and measurement of terms of trade	359		
Causes of changes in the terms of trade	361		
Consequences of changes in the terms of trade	363		
4.1 Economic Development	365		
Economic growth and economic development	365		
Common characteristics of and diversity among economically less developed countries	367		
International development goals	373		
4.2 Measuring Economic Development	374		
Single indicators of development	374		
Composite indicators of development	382		
4.3 The Role of Domestic Factors in Economic Development	386		
Domestic factors and economic development	386		
4.4 The Role of International Trade in Economic Development	397		
Trade problems facing many LEDCs	397		
Trade strategies for economic growth and economic development	402		

the



Economics

classroom

Mastery learning tools for teachers and students

<http://econclassroom.com>

jason@econclassroom.com

follow on:

[youtube](#)

[twitter](#)

[facebook](#)

[linkedin](#)

Introduction to the Series

Thank you for purchasing the Economics Classroom IB Economics class notes. This resource is just one of many available from The Economics Classroom, a website created in 2007 to provide economics students and teachers with resources to help them master the subject in an easy and enjoyable manner.

Jason Welker taught International Baccalaureate and Advanced Placement Economics at three international schools in Asia and Europe over a 13 year career before returning to his home in the United States to focus on creating resources for economics students and teachers full time. After a collaboration with Khan Academy on their AP Economics content, he turned his attention to producing new and better resources for AP and IB students and making them available through EconClassroom.com.

If you find these notes useful, you may want to sign up for some of the other services available through EconClassroom.com. These include tutoring, exam revision ,support for written work (like the internal assessment and extended essay), and a subscription option that provides access to the great content in these notes and all the videos from Jason's YouTube channel (www.youtube.com/jasonwelker), practice activities for every lesson, and complete practice exams to help you prepare for the AP Micro and Macroeconomics exams.

Thanks again for purchasing this resource. Please spread the word among your classmates and teachers about how useful it is for you. I look forward to providing you with more support as you work towards mastering economics in the coming months!

- Jason Welker



1.0 Introduction to Economics

Economics as a social science

- Explain that Economics is a social science
- Outline the social scientific method.
- Explain the process of model building in economics.
- Explain that economists must use the *ceteris paribus* assumption when developing economic models.
- Distinguish between positive and normative economics.
- Examine the assumption of rational economic decision-making

Economics is the **social science** that studies the interactions of humans in the commercial realm. Economists examine the way societies allocate their scarce resources towards competing wants and needs and seek to develop systems that achieve certain objectives, including:

- Growth in humans' standard of living over time
- Sustainable development
- Employment and stability

Microeconomics vs. Macroeconomics

Microeconomics studies the behaviors of individuals within an economy, including consumers and producers in the market for particular goods. Examples of microeconomic topics:

- the Automobile market in Tennessee,
- the market for movie tickets in Seattle,
- the market for airline tickets between the US and Europe,
- the market for vacations to Spain,
- the market for teachers

Macroeconomics studies the total effect on a nation's people of all the economic activity within that nation. The four main concerns of macroeconomics are:

- total output of a nation,
- the average price level of a nation,
- the level of employment (or unemployment) in the nation,
- the distribution of income in the nation

Examples of macroeconomic topics include: unemployment in Canada, inflation in Zimbabwe, economic growth in China, the gap between the rich and the poor in America.

Microeconomics and macroeconomics can be broken down into many smaller topics. Some of them are identified below.

Microeconomics topics	Macroeconomics topics
Individual markets the behavior of firms (companies) and consumers the allocation of land, labor and capital resources Supply and demand The efficiency of markets Product markets Supply and Demand Profit maximization Utility maximization Competition Market failure	National markets Total output and income of nations Total supply and demand of the nation Taxes and government spending Interest rates and central banks Unemployment and inflation Income distribution Economics growth and development International trade

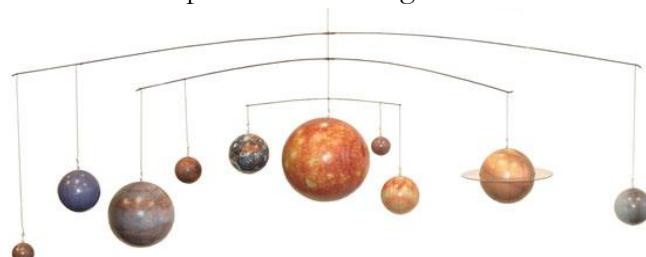
Fundamental concepts of economics

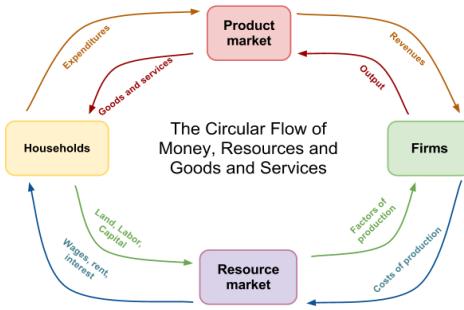
Whether we study micro or macro, there are some basic concepts that underlie all fields of economics:

Scarcity:	Economics is about the allocation of scarce resources between society's various needs and wants.
Resources:	Economics is about the allocation of resources among society's various needs and wants.
Tradeoffs:	Individuals and society as whole are constantly making choices involving tradeoffs between alternatives. Whether it's what goods to consume, what goods to produce, how to produce them, and so on.
Opportunity Cost:	“The opportunity cost is the opportunity lost”. In other words, every economic decision involves giving up something. NOTHING IS FREE!!

Model building in economics

A popular tool in the Economist's kit is the **economic model**. Just like scientists in other fields, economists use models to represent something from the real world.





A model of the solar system allows astronomers to illustrate in a simplified model the relationships between solar bodies.

A Circular Flow Model allows economists to illustrate in a simplified model of the relationships between households and firms in a market economy.

Ceteris Paribus is a Latin phrase meaning “all else equal”. As in other sciences, when using economic models economists must assume “all other variables are held constant.” This allows economists to observe how one variable in an economy will affect another, without considering all the other factors that may affect the variable in question.

Positive and normative economics

Economists explore the world of facts and data, but also often draw conclusions or prescribe policies based more on interpretation or even their own opinions. It is important to distinguish at all times whether the focus of our studies is in the realm of positive or normative Economics

Positive economics deals with statements of fact. Positive economics can be supported by evidence based on quantifiable observations of the world. Examples of positive economics statements include:

- Unemployment rose by 0.8 percent last quarter as 250,000 Americans lost their jobs in both the public and private sectors.
- Rising pork prices have led to a surge in demand for chicken across China.
- Increased use of public transportation reduces congestion on city streets and lowers traffic fatality rates.

Normative economic statements are based on observable, quantifiable variables, but include an element of opinion.

- Unemployment rates are higher among less educated workers, therefore government should include education and job training programs as a component of benefits for the nation's unemployed.
- Rising pork prices harm low-income households whose incomes go primarily towards food, therefore, to slow the rise in food prices, the Chinese government should enforce a maximum price scheme on the nation's pork industry.
- It is the government's obligation to provide public transportation options to the nation's people to relieve the negative environmental and health effects of traffic

congestion.

Scarcity

- Explain that scarcity exists because factors of production are finite and wants are infinite
- Explain that economics studies the ways in which resources are allocated to meet needs and wants
- Explain that the three basic economic questions that must be answered by any economic system are: “What to produce?”, “How to produce?” and “For whom to produce?”

Introduction to scarcity

You may not know it yet, but you are beginning a science class. Yes, Economics is a science, and just like other sciences, it deals with a fundamental problem of nature.

- Think of Aerospace Engineering. This is a science that struggles to overcome a basic problem of nature, that of gravity. Aerospace Engineers are scientists whose research and life's work is aimed at overcoming the problem of gravity and putting man in space.
- Economists are also scientists whose work attempts to overcome a basic problem of nature.

A riddle:

What is the basic problem of nature that the science of Economic attempts to overcome?

Hint: It arises because of the limited nature of earth's natural resources!

The answer is **scarcity**! Scarcity is the basic problem of Economics. Scarcity arises when something is both limited in quantity yet desired

Some facts about scarcity

- Not all goods are scarce, but most are
- Some goods that humans consume are infinite, such as air

Scarce (limited and desired)	Not Scarce (not limited OR not desired)
diamonds, apartments, drinking water, teachers, doctors, cars, medical services	air, salt water, mosquitos, malaria, love, HIV , crime, established knowledge

So, what makes something scarce? Here's another riddle for you...

- Nobody needs diamonds, yet they are extremely valuable
- Everybody needs water, yet they are extremely cheap

Why are diamonds so expensive? Why is water so cheap? This is known as the “diamond/water paradox”. The answer lies in the fact that economic value is derived from scarcity

- The more scarce an item, the more valuable it is

- The less scarce, the less value it has in society!

Free goods and economics goods

Goods in economics are those things we like to consume. They are called “goods” because consuming them makes us feel good!

- **Free goods** are those things that we desire but that are not limited, or which are limited but no one desires
- **Economic goods** are those that we desire but that ARE limited

Economic Goods (limited and desired)	Free Goods (unlimited or undesired)
Haircuts, shoes, education, cars, vacations, toothbrushes, fresh drinking water, hamburgers, televisions, public transportation, movies, jewelry	Air, friendship, salt water, love, clouds, happiness, established knowledge, disease, violence, conflict

The factors of production

The production of all of the goods we desire requires **scarce resources**. Scarce resources are also known as the **factors of production**. It is the allocation of the factors of production between humans' competing wants on which the social science of economics focuses.

The Factors of Production (resources)			
Land	Labor	Capital	Entrepreneurship
Land resources are those things that are "gifts of nature". The soil in which we grow food, wood, minerals such as copper and tin and resources such as oil, coal, gas and uranium are scarce	Labor refers to the human resources used in the production of goods and services. Labor is the human work, both physical and intellectual, that contributes to the production of goods and services	Capital refers to the tools and technologies that are used to produce the goods and services we desire. Since more and better tools enhance the production of all types of goods and services, from cars to computers to education to haircuts, yet the amount of capital in the world is limited, capital is a scarce resource.	Entrepreneurship refers to the innovation and creativity applied in the production of goods and services. The physical scarcity of land, labor and capital does not apply to human ingenuity, which itself is a resource that goes into the production of output economic output.

“The Basic Economic Problem”

In a world of finite resources, the wants of man are virtually infinite. The **basic economic problem** is how to allocate those limited, scarce resources between the unlimited wants of man. This problem gives rise to three questions that any and all economic systems must address. The **three basic economics questions** are :

1. **What should be produced?** Given the resources with which society is endowed, what combination of different goods and services should be produced?
2. **How should things be produced?** Should production use lots of labor, or should lots of capital and technology be used?

3. **Who should things be produced for?** How should the output that society produces be distributed? Should everyone keep what he or she makes, or should trade take place? Should everyone be given equal amounts of the output, or should it be every man for himself?

These are the three guiding questions of any economic system. **Economic systems** are the systems society adopts for answering the three economic questions above. A society's actual system for resource allocation falls along a spectrum from free markets to command economies.

The free market system: Free market economies are those in which government has no control over the ownership or allocation of resources. What gets produced, how it is produced, and who gets it are all determined by the interactions of private individuals engaging in voluntary trade in markets: places where buyers and sellers meet to trade goods, services, and factors of production.

The command system: Also called “communism” or “socialism”, a command economy is one in which a central planning authority (usually a nation’s government) owns and controls the allocation of all the factors of production. What gets produced, how it is produced, and who gets it are all determined by the government. The aim of most command economies throughout history has been to eliminate class differences and promote total equality in how resources and production is allocated.

The mixed economic system: In reality, nearly every society today employs an economic system with characteristics of both the market and command systems. While most goods and services are produced by private businesses and consumed by households who are free to decide what to buy and what to do with their own privately owned resources (labor, capital, and land), some goods and services are also provided by government. Think of the roads you drive on, the school you attend, and the protection you enjoy by the police and military. These “public goods” are provided by government, while most of the things you buy are provided by the market. Mixed economies characterize nearly every country’s system of resource allocation in the 21st century.

Choice and Opportunity Cost

- Explain that as a result of scarcity, choices have to be made
- Explain that when an economic choice is made, an alternative is always foregone
- Explain that a production possibilities curve (production possibilities frontier) model may be used to show the concepts of scarcity, choice, opportunity cost and a situation of unemployed resources and inefficiency.

Because resources are scarce, society faces tradeoffs in how to allocate them between different uses. Every choice about the use of a resource comes with an opportunity cost, and these choices can be illustrated in a simple model called the **Production Possibilities Curve (PPC)**.

Opportunity cost

Perhaps the most fundamental concept to economics, **opportunity cost** is what must be given up in order to undertake any activity or economic exchange.

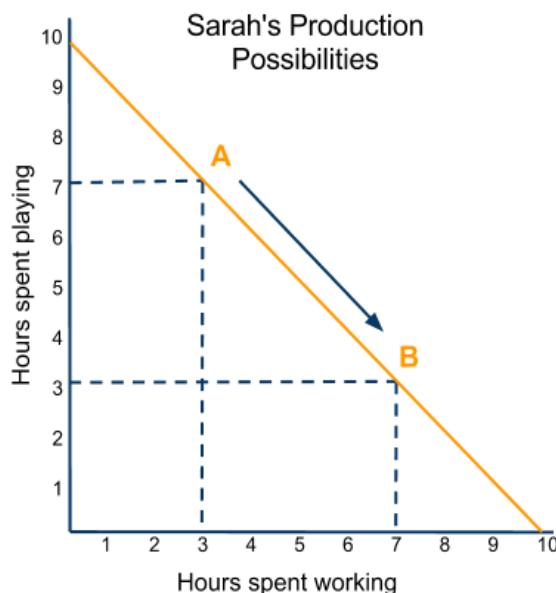
- Opportunity costs are not necessarily monetary, rather when you buy something, the opportunity cost is what you could have done with the money you spent on that thing.
- Even non-monetary exchanges involve opportunity costs, as you may have done something different with the time you chose to spend undertaking any activity in your life.

Examples of opportunity costs

- The opportunity cost of watching TV on a weeknight is the benefit you could have gotten from studying.
- The opportunity cost of going to college is the income you could have earned by getting a job out of high school
- The opportunity cost of starting your own business in the wages you give up by working for another company
- The opportunity cost of using forest resources to build houses is the enjoyment people get from having pristine forests.

The Production Possibilities Curve

The tradeoff we face between the use of our scarce resources (or even time) can be modeled in a simple economic graph known as the Production Possibilities Curve (the PPC). The PPC here shows how Sarah can use her limited free time of 10 hours per day to either “work” or “play”.



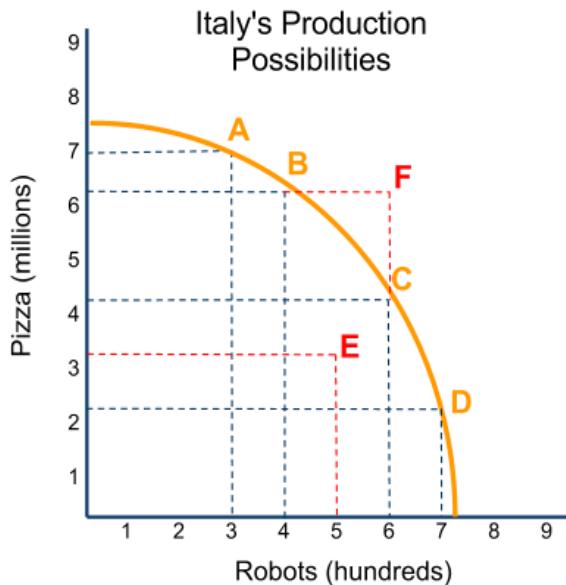
Sarah faces two **tradeoffs**. She can either work or play with her limited amount of time.

- The opportunity cost of an hour of work is an hour of play
- As she goes from 3 hours of work to 7 hours of work, she gives up 4 hours of play.

- She cannot spend 10 hours working AND 10 hours playing, so Sarah has to make choices.

The PPC model can be used to illustrate the economic challenges faced by individuals, firms, states, countries or the entire world.

Consider the hypothetical PPC for the country of Italy.



This PPC shows that Italy can produce:

- Either 7.5 million pizzas,
- OR 750 robots
- Note, however, that Italy can NOT produce 7.5 million pizzas AND 750 robots

Italy faces a tradeoff in how to use its scarce resources of land, labor and capital. As the country moves along its PPC from point A to point D:

- It gives up more and more pizza to have more robots
- It gives up current consumption of food for production of robots, which themselves are capital goods, and therefore will assure that Italy's economy will grow into the future.

Assumptions about the PPC

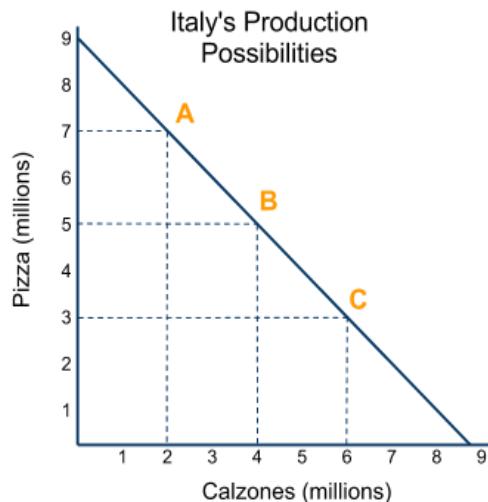
- A point ON the PPC is attainable only if a nation achieves full-employment of its productive resources
- The nation's resources are fixed in quantity
- The economy is closed, i.e. does not trade with other countries
- Represents only one country's economy

Observations about points on or within the PPC

- Points on the PPC are attainable, and desirable, since a country producing on the line

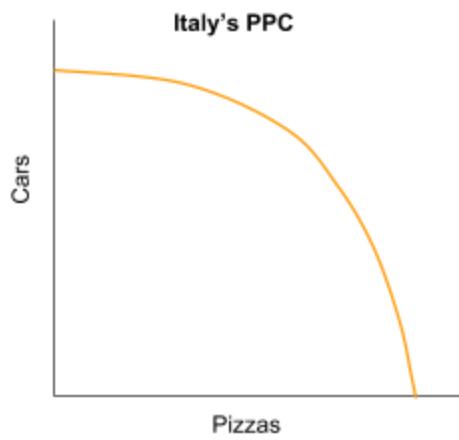
- is achieving full employment and efficiency
- Points inside the PPC (such as E) are attainable but undesirable, because a nation producing here has unemployment and is inefficient
 - Points outside the PPC (such as F) are unattainable because they are beyond what is presently possible given the country's scarce resources. But such points are desirable because they mean more output and consumption of both goods.

A PPC can be either straight or bowed outwards from the origin



A straight line PPC:

- Indicates that the two goods require similar resources to produce (like pizzas and calzones)
- The opportunity cost of one pizza is one calzone, so Italy always gives up the same quantity of one good no matter where it is on its PPC

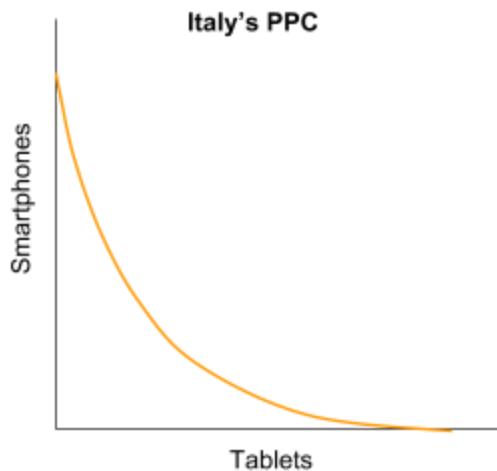


A bowed out PPC:

- Indicates that the two goods require very different resources to produce (like pizzas and cars)
- As Italy increases its output of one good, the opportunity cost (in terms of the

quantity of the other good that must be given up) increases. This illustrates the law of increasing opportunity cost

The **law of increasing opportunity cost** says that as the output of one good increases, the opportunity cost in terms of other goods tends to increase.



A PPC that is bowed inward

- Indicates that as the output of one good increases, the opportunity cost of (in terms of the quantity of the other good that must be given up) decreases.
- Indicates decreasing opportunity cost. Only likely if the resources needed to produce one good become less scarce as the production of the other good increases.

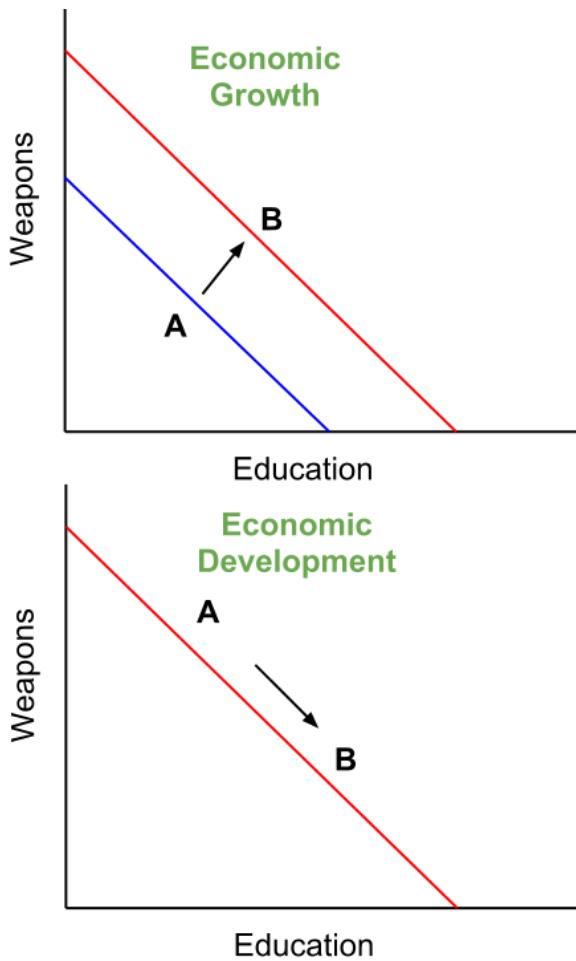
Key concepts shown by the PPC

In addition to opportunity costs and tradeoffs, the PPC can be used to illustrate several other key Economic concepts, including...

- **Scarcity:** Because of scarcity, society can only have a certain amount of output
- **Actual output:** A country's actual output is shown by where it is currently producing on or within its PPC
- **Potential output:** A point on the PPC shows the potential output of a nation at a particular time
- **Economic growth:** An increase in the quantity or the quality of a nation's resources will shift its PPC out, indicating the economy has grown
- **Economic development:** The composition of a nation's output will help determine whether the standards of living of its people are improving over time

The distinction between economic growth and economic development

Economic growth refers to the increase in the total output of goods and services by a nation over time.



- It is also sometimes defined as an increase in household income over time.
- It is purely a monetary measure of the increases in the material well being of a nation.
- On a PPC growth can be shown as an outward shift of the curve.

Economic development refers to the improvement in people's' standard of living over time.

- Measured by improvements in health, education, equality, life expectancy and so on
- Incorporate income as well, but is a much broader measure than growth
- On a PPC development can be shown by a movement towards the production of goods that improve people's' lives

Central Themes

- The extent to which governments should intervene in the allocation of resources
- The threat to sustainability as a result of the current patterns of resource allocation
- The extent to which the goal of economic efficiency may conflict with the goal of equity
- The distinction between economic growth and economic development

Other central themes to the IB economics course include:

- **The role of government in the economy:** In every unit of this course we will examine the appropriate role of government in the market economy. There are some who argue government should never interfere with the free functioning of markets; on the other hand, when market failures arise, the government may be needed to improve the allocation of resources.
- **Threats to sustainability of current economic trends:** What threat do global warming, environmental degradation, population growth and urbanization play to the ability of our economic systems to endure?
- **The conflict between the pursuits of efficiency and equity:** Sometimes the pursuit of wealth and economic growth leaves some individuals behind. To what extent should economic policy be concerned with income and wealth inequality? Is there a mechanism available for reducing inequality while at the same time encouraging efficiency?

1.1 Competitive Markets, Demand and Supply

Markets

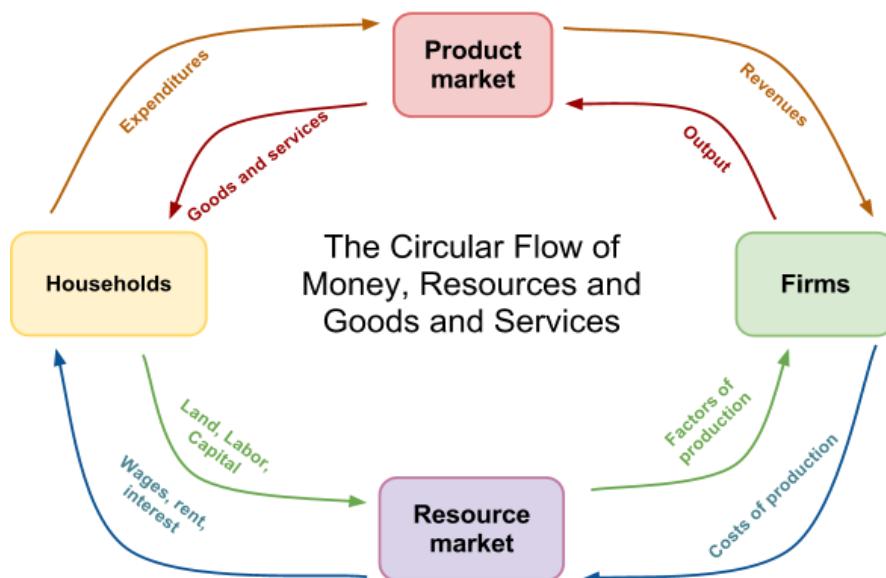
- The natures of markets
- Outline the meaning of the term market

Recall from the previous chapter that **the market system** is that which most economies today are based on. Markets come in many forms, but most can be characterized as one of the following

Type	Resource Market	Product Market
What gets bought and sold?	Land, Labor, Capital and Entrepreneurship	Goods and services
Who are the demanders?	Business firms demand resources	Households
Who are the suppliers?	Households supply resources	Firms supply products made with the resources provided by households
Money flows...	From firms to households as wages, interest, rent and profits	From households to firms as expenditures (revenues for firms)
Examples	The market for: bus drivers, waitresses, bankers, janitors	The markets for: bus journeys, restaurant meals, financial services, cleaning services

Markets in the circular flow model

The **circular flow model** shows the flow of money payments between households and firms in the market economy.



The circular flow model represents how an economy with a well defined system of property rights in which individuals respond to incentives in pursuit of certain objective functions.

Some key takeaways from the model of the market system are:

- Households and firms are interdependent on one another.
- Households pursue happiness through consumption
- Firms pursue profits through the sale of their products to households
- All exchanges are voluntary and mutually beneficial: households earn income in the resource market and firms earn revenues in the product market.

All markets have two sides: the Demand side and the Supply side. We'll now examine what is meant by "demand", how it is determined, and what can cause demand to increase or decrease in a particular market.

The law of demand

- Explain the negative causal relationship between price and quantity demanded.
- Describe the relationship between an individual consumer's demand and market demand.

Introduction to demand

In order for a market to function, there must be **demand** for a product or a resource. But what, exactly IS demand?

Think of your favorite candy, and ask yourself, how much of it would you be willing to buy in one week if it cost the following: \$5, \$4, \$3, \$2, \$1.

The table below shows how many Butterfingers™ Carl would buy in a week at each of the prices indicated.

Price	Quantity demanded
\$5	1
\$4	2
\$3	3
\$2	4
\$1	5

Notice that as the price of Butterfingers decreases, Carl demands a greater quantity. This is Carl's individual demand for Butterfingers.

Demand is defined as the quantity of a particular good that consumers are willing and able to buy at a range of prices at a particular period of time.

From individual demand to market demand

The table above represents only Carl's demand. But a market is made up of many consumers and producers. Market demand is the sum of the individual demands of all the consumers in the market.

For example, assume that Carl is one of 5 individual consumers who demand Butterfingers in a small town (a really small town!)

The table below shows the weekly individual demands of the five consumers for Butterfingers at a range of prices.

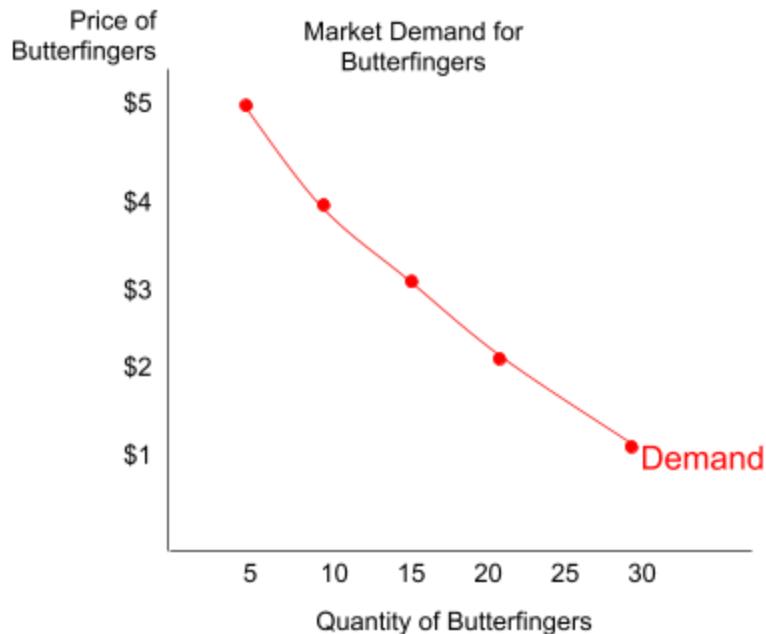
Price	Carl's quantity	Sarah's quantity	Henry's quantity	Lara's quantity	Jade's quantity	Total Demand
\$5	1	0	3	0	1	5
\$4	2	1	3	1	2	9
\$3	3	1	4	3	3	14
\$2	4	1	6	5	6	21
\$1	5	2	8	7	7	29

This is the **market demand** for candy in a week. The market demand is simply the sum of all the individual consumers' demands in a market

The demand curve

- Explain that a demand curve represents the relationship between the price and the quantity demanded of a product, ceteris paribus.
- Draw a demand curve.

The table representing how much candy would be demanded at a range of prices is called a demand schedule. This data can be represented graphically in a **demand curve**. The graph below shows the market demand for candy in our small town.



The demand curve plots the quantity demanded along the horizontal axis against the price on the vertical axis. The red dots represent a price/quantity combination from the demand schedule. By connecting the dots we get a demand curve.

Some observations about the demand curve:

- At higher prices, a smaller quantity is demanded
- At lower prices, a larger quantity is demanded
- The curve is downward sloping
- There is an inverse relationship between price and quantity demanded

The law of demand

The downward sloping demand curve illustrates the **law of demand**, which states that ceteris paribus (all else equal), there is an inverse relationship between a good's price and the quantity demanded by consumers

While it may seem obvious that people want to buy more stuff at lower prices and less stuff at higher prices, there are a few explanations economists have developed for the law of demand. These are the income effect, the substitution effect, and the law of diminishing marginal utility.

The income effect observes that as the price of a good decreases, the quantity demanded increases because consumers now have more real income to spend. Real income refers to income that is adjusted for price changes, and implies the actual buying power of a consumer. With more buying power, they sometimes choose to buy more of the same product.

The substitution effect observes that as the price of a good decreases,

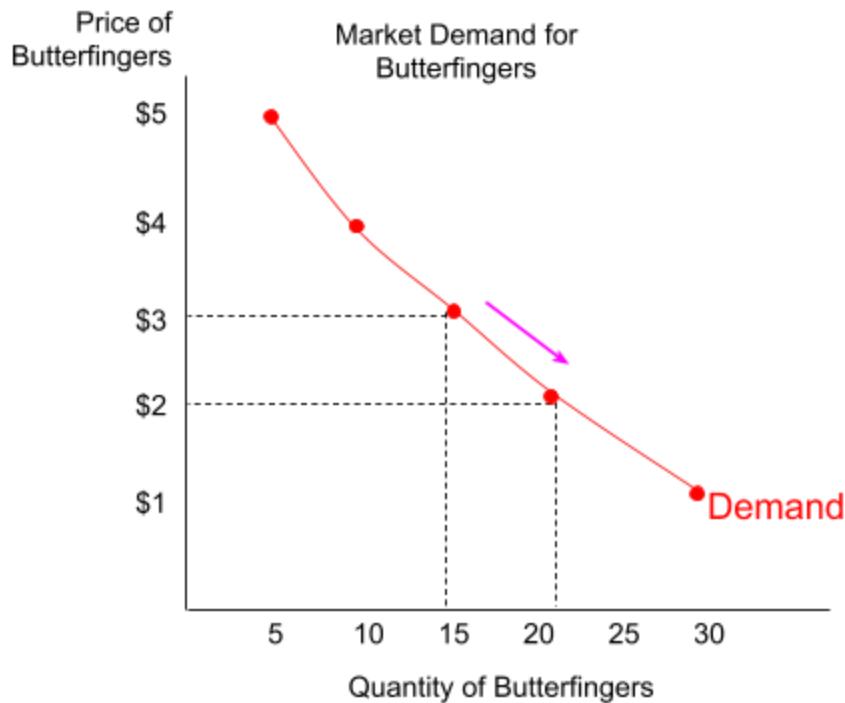
consumers switch from other substitute goods to this good because its price is comparatively lower.

The law of diminishing marginal utility states that as we consume additional units of something, the satisfaction (utility) we derive for each additional unit (marginal unit) grows smaller (diminishes). Therefore, consumers are only willing to buy more of a good when its price falls.

Changes in demand vs. changes in quantity demanded

There is a difference between a **change in demand** and a **change in quantity demanded**. If all other factors are held constant and only a good's price changes, there is a change in the quantity demanded for the good.

For example, if actual price of Butterfingers falls from \$3 to \$2, there will be a movement along the demand curve and a change in quantity demanded, as seen in the graph below.



A change in price leads to a change in the quantity demanded

- As seen in the graph, when the price of Butterfingers falls, a greater quantity is demanded.
- When the price of candy rises, a smaller quantity is demanded.
- A change in price leads to a change in the quantity demanded.

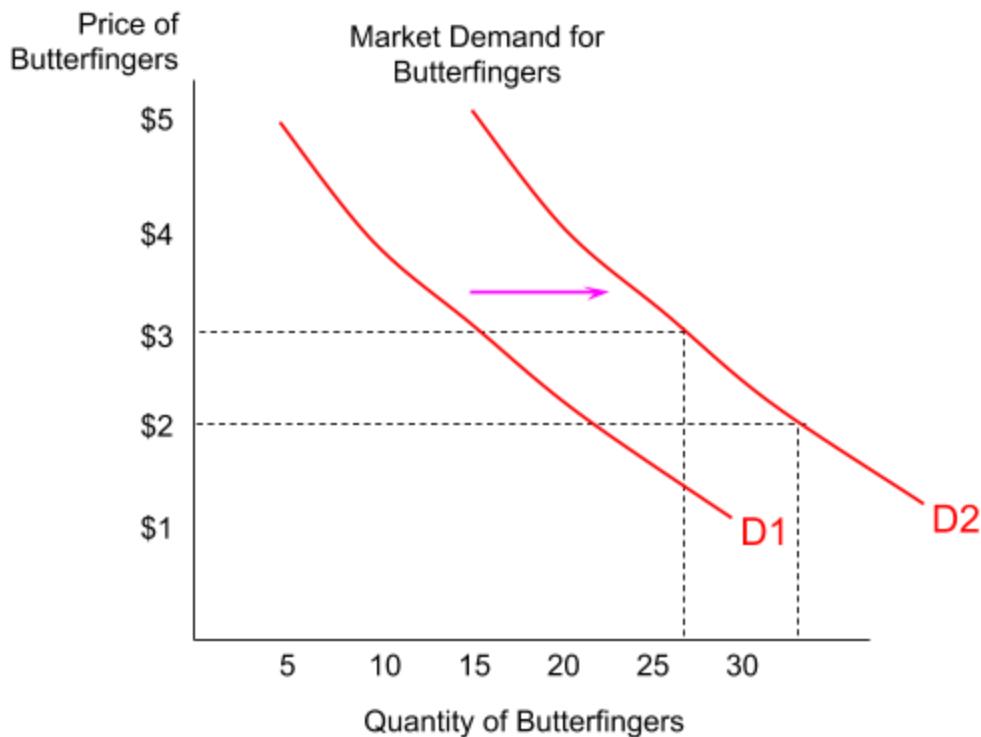
The non-price determinants of demand and movements along vs. shifts of the demand curve

- Explain how factors including changes in income (in the cases of normal and inferior

goods), preferences, prices of related goods (in the cases of substitutes and complements) and demographic changes may change demand.

- Distinguish between movements along the demand curve and shifts of the demand curve.
- Draw diagrams to show the difference between movements along the demand curve and shifts of the demand curve.

Now let's assume that a factor other than the price of Butterfingers changes. For example, assume the five consumers in this market experience an increase in income and as a result they demand more Butterfingers at every price. The entire demand curve will now shift out as seen in the graph below.

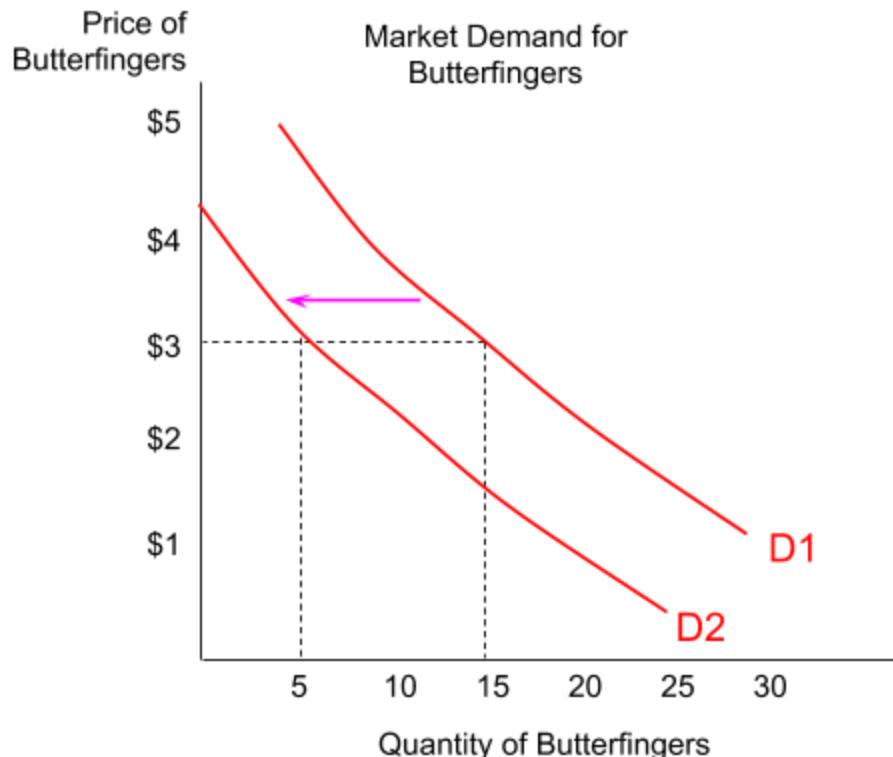


At the new level of demand “D2” more candy bars are demanded at every price. This “increase in demand” resulted from an increase in consumers incomes. Besides incomes, there are several other **determinants of demand**, a change in which will cause the entire demand curve for a good to shift outwards (increase) or inwards (decrease).

The Determinants of Demand (Demand shifters)	
Tastes and preferences of consumers	A change in consumers' tastes and preferences will cause demand for a good to either increase (if it is more preferred) or decrease (if it is less preferred)
Other related goods' prices	If a substitute good (for example, Snickers bars) gets cheaper, the demand for the good in question will fall. If substitutes' price increase, demand increase for

	the good in question.
	If a complementary good (one that is usually consumed with the good in question) gets cheaper, demand for a good will increase. If complements' prices increase, demand for the good in question decreases.
Expectations of future prices	If consumers expect a good's price to rise in the future, demand today will increase. If prices are expected to fall, demand will decrease today as consumers postpone their consumption until the future.
Incomes	For normal goods, higher incomes will increase demand and lower incomes will decrease demand. However, if a good is "inferior" then a fall in incomes will increase demand and an increase in income will decrease demand.
Size of the market	An increase in the number of consumers will increase the demand for a good; a decrease in the number of consumers causes demand to fall.
Special circumstances	Changes in factors such as weather, natural disasters, scientific studies, and so on, will cause demand for some goods to change. For instance, an unusually heavy winter may increase demand for ski resorts, as more people head to the mountains to enjoy the snow!

Let's look at one example of how a change in a determinant of demand for Butterfingers can affect the market demand for the candy bar. Assume that the price of Snickers, another popular candy, falls. The graph below shows the effect this will have on demand for Butterfingers.



Due to a decrease in the price of a close substitute, the demand for Butterfingers has

decreased from D1 to D2. Now, instead of 14 Butterfingers being demanded at a price of \$3, only 5 are demanded. Consumers have switched to the now cheaper substitute, and demand has decreased for Butterfingers.

Linear demand equations, demand schedules and graphs (HL only)

- Explain a demand function (equation) of the form $Q_d = a - bP$.
- Plot a demand curve from a linear function (eg. $Q_d = 60 - 5P$).
- Identify the slope of the demand curve as the slope of the demand function $Q_d = a - bP$, that is $-b$ (the coefficient of P).
- Outline why, if the “ a ” term changes, there will be a shift of the demand curve.
- Outline how a change in “ b ” affects the steepness of the demand curve.

Demand, which we have now seen expressed in both a schedule and as a curve on a diagram, can also be expressed mathematically as an equation. We will examine linear demand equations, which are simple formulas that tell us the quantity demanded for a good as a function of the good's price and non-price determinants.

A typical demand equation will be in the form:

$$Q_d = a - bP$$

Where:

- ‘ Q_d ’ = the quantity demanded for a particular good
- ‘ a ’ = the quantity demanded at a price of zero. This is the ‘ q -intercept’ of demand, or where the demand curve crosses the Q -axis
- ‘ b ’ = the amount by which quantity will change as price changes, and
- ‘ P ’ = the price of the good

Consider the demand for bread in a small village, which can be represented by the following equation:

$$Q_d = 600 - 50P$$

What do we know about the demand for bread from this function? We know that:

- If bread were free (e.g. if the price = 0), 600 loaves of bread would be demanded. Plug zero into the equation to prove that $Q_d = 600$
- For every \$1 increase in the price of bread above zero, 50 fewer loaves will be demanded. We can plug the following prices into the equation to prove this:

$$\begin{aligned} \$1 - Q_d &= 600 - 50(1) = 550 \\ \$2 - Q_d &= 600 - 50(2) = 500 \\ \$3 - Q_d &= 600 - 50(3) = 450 \\ \$4 - Q_d &= 600 - 50(4) = 400 \end{aligned}$$

We can also calculate the price at which the quantity demanded will equal zero. This is known as **the price-intercept** (because it's where the demand curve crosses the price axis).

To prove this, set Q equal to zero and solve for P.

$$0=600-50(P).$$

$$P = 12$$

Creating a demand schedule from a demand equation

A demand equation can be plotted in both a demand schedule and as a demand curve. In the market for bread, we already determined the following:

- At a price of \$0, the quantity demanded is 600 loaves. This is the q-intercept
- At a price of \$12, the quantity demanded is 0 loaves. This is the p-intercept

With these numbers, we can create a demand schedule for the demand equation

$$Q_d=600-50P$$

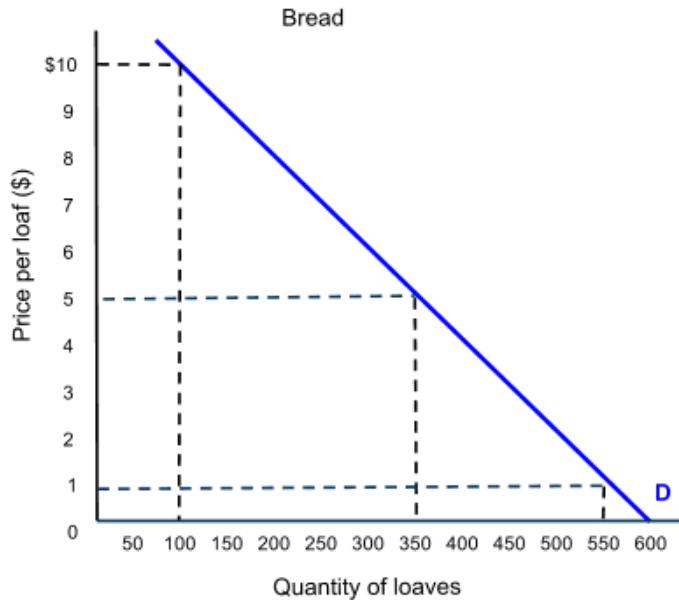
Price per loaf	Quantity of loaves demanded
0	600
2	500
4	400
6	300
8	200
10	100
12	0

Notice that for every \$2 increase in the price, the quantity demanded falls by 100 loaves. This corresponds with our 'b' variable of 50, which tells us how responsive consumers are to price changes. For every \$1 increase in price, 50 fewer loaves are demanded.

Plotting a demand curve from a demand equation

The data from our demand schedule can easily be plotted on a graph. OR, we could have just plotted the two points of demand we knew before creating the demand schedule.

- The Q-intercept of 600 loaves, and
- The P-intercept of \$12



Notice the following:

- The demand for bread is inversely related to the price. This reflects the law of demand
- The slope of the curve is negative, this is reflected in the equation by the ‘-’ sign in front of the ‘b’ variable.
- For every \$1 increase in price, Q_d decreases by 50 loaves.
- 50 is NOT the slope of demand, however, rather, it is the ‘run over rise’. In other words, the ‘b’ variable tells us the change in quantity resulting from a particular change in price.

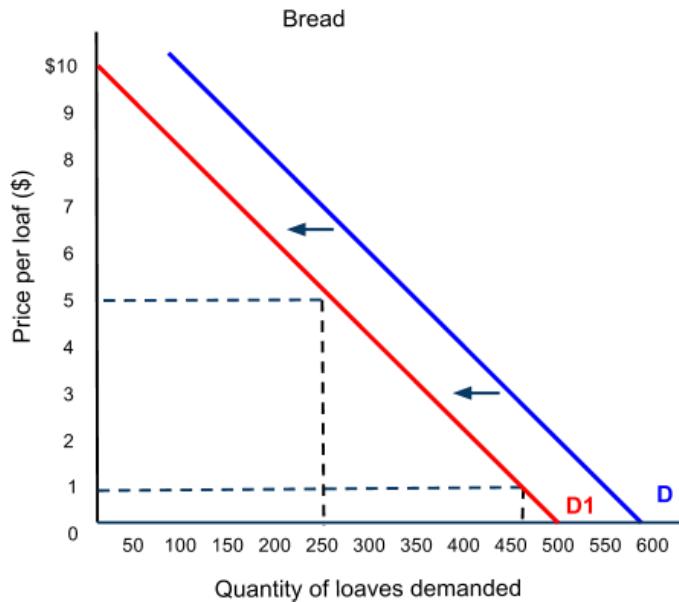
Changes in the ‘a’ variable

As we learned earlier, a change in price causes a change in the quantity demanded. This relationship can clearly be seen in the graph.

- But what could cause a shift in the demand curve?
- And how does this affect the demand equation?

A change in a non-price determinant of demand will change the ‘a’ variable.

- Assume the price of rice, a substitute for bread, falls.
- Demand for bread will decrease and the demand curve will shift.
- In the demand equation, this causes the ‘a’ variable to decrease. Assume the new equation is: $Q_d=500-50P$



Following the change in the 'a' variable, less bread will be demanded at every price. The new Q-intercept is only 500 loaves. The demand curve has shifted to the left.

Notice the following:

- At each price, 100 fewer loaves are now demanded. In the original graph, 350 loaves were demanded at \$5, now only 250 are demanded.
- Demand has decreased because a non-price determinant of demand changed (the price of a substitute decreased, so consumers switched to rice).
- The 'b' variable did not change, so the slope of the demand curve remained the same.
- The P-intercept decreased to \$10. Now, at a price of \$10, no bread is demanded, whereas before consumers would buy bread up to \$12.

Changes in the 'b' variable

The 'b' variable in the demand equation is an indicator of the responsiveness of consumers to price changes.

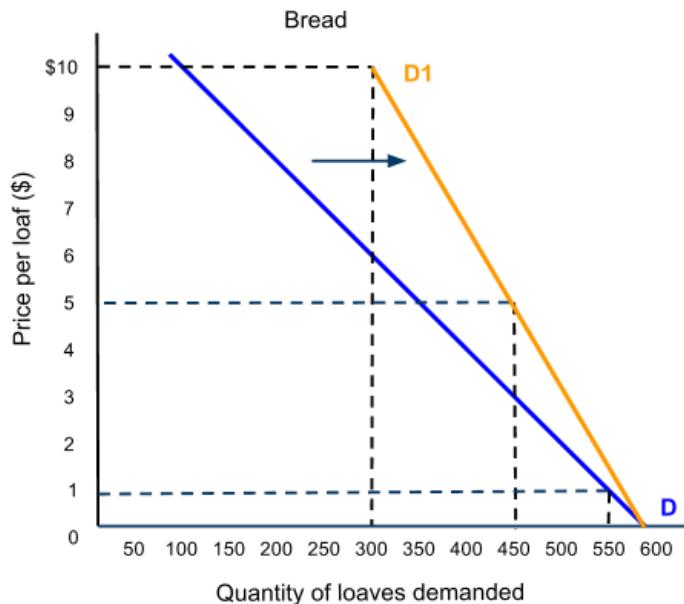
- If something causes consumers to be more responsive to price changes, the 'b' variable will increase
- If something causes consumers to be less responsive to price changes, the 'b' variable will decrease

Assume several bakeries have shut down in the village and only one remains. Consumers now have less choice and must buy their bread from that bakery, therefore they become less responsive to price changes. The 'b' variable in the equation will decrease to 30

$$Q_d = 600 - 30P$$

Now, for every \$1 increase in price, consumers will demand 30 fewer loaves, instead of 50.

The Q-intercept will remain the same (600) but the demand curve will be steeper, indicating consumers are less responsive to price changes



Notice the following:

- Consumers are less responsive to price changes now.
- As the price rises from \$0 to \$5 per loaf, now consumers will still demand 450 loaves, whereas in the original graph they would have only demanded 350 loaves.
- Demand for bread has increased because there are fewer substitutes in this village.
- The new P-intercept is not visible on the graph, but it can easily be calculated. Set Q to zero and solve for P

$$\begin{aligned} 0 &= 600 - 30P \\ P &= 20 \end{aligned}$$

Now, at a price of \$20, zero loaves will be demanded

The law of supply

- Explain the positive causal relationship between price and quantity supplied.
- Describe the relationship between an individual producer's supply and market supply

All markets include buyers and sellers. The buyers in a market demand the product, but the sellers supply it. Let's consider the market for candy bars. The table below shows how many candy bars would be supplied weekly by an individual retailer (a convenient store) in a small town at a range of prices.

Price	Quantity supplied
\$5	100
\$4	80

\$3	60
\$2	40
\$1	20

Notice that at higher prices a greater quantity of candy bars would be supplied. At lower prices, fewer candy bars are supplied.

The **individual supply** schedule above represents an individual seller's supply. **Market supply** is made up of the sum of all the individual sellers in a market. The table below sums up the individual weekly supplies of candy bars of the five convenient stores in a small town.

Price	7/11's quantity	Quick Stop's quantity	Thrifty's quantity	Viva Express's quantity	Gas'n'Go's quantity	Total Supply
\$5	100	40	60	70	30	300
\$4	80	30	45	60	25	240
\$3	60	20	30	45	20	165
\$2	40	10	15	30	15	110
\$1	20	5	5	15	10	55

This is the market supply of candy bars in a week in the small town. Notice that market supply is the horizontal sum of the individual supply schedules of the sellers in a market.

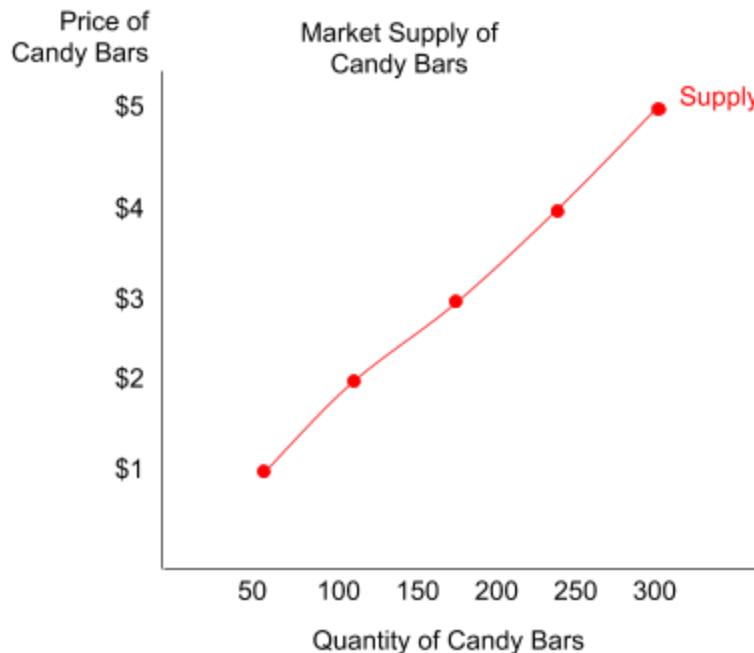
Supply is defined as a schedule or curve showing how much of a product producers are will supply at each of a range of possible prices during a specific time period.

The **law of supply** states that *ceteris paribus* (all else equal), there is a direct relationship between the price of a product and the quantity supplied. As the price of a good increase, firms will increase their output of the good. As price decreases, firms will decrease their output of the good.

The supply curve

- Explain that a supply curve represents the relationship between the price and the quantity supplied of a product, *ceteris paribus*.
- Draw a supply curve

The data from a supply schedule can be presented in a graph using the same axes and labels as the demand graph. For example, the market supply of candy bars can be plotted, as in the graph here.



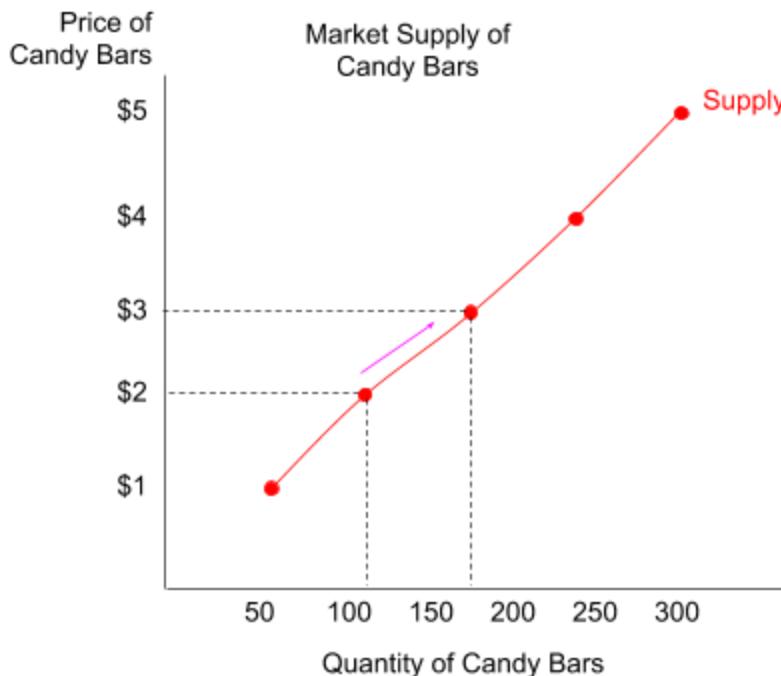
Consider the market for candy bars:

- An increase in the price of candy results in more candy being produced, as more firms can cover their costs and existing firms increase output.
- A fall in the price of candy results in the quantity supplied falling, as fewer firms can cover their costs, they will cut back production.
- Only the most efficient firms will produce candy at low prices, but at higher prices more firms enter the market

The non-price determinants of supply and movements along vs. shifts of the supply curve

- Explain how factors including changes in costs of factors of production (land, labour, capital and entrepreneurship), technology, prices of related goods (joint/competitive supply), expectations, indirect taxes and subsidies and the number of firms in the market can change supply.
- Distinguish between movements along the supply curve and shifts of the supply curve.
- Draw diagrams to show the difference between movements along the supply curve and shifts of the supply curve.

When a good's price changes, there is a **change in the quantity supplied** of the good and a movement along the supply curve. Assume, for example, the price of candy bars increases from \$2 to \$3, as in the graph below.



As price increases, the quantity of candy bars retailers are willing to supply increases. There is a movement up and to the right along the supply curve. A decrease in price would cause the quantity supplied to decrease and a movement down and to the left along the supply curve.

A change in a factor other than the price will cause a shift in the supply curve, and a **change in supply** (shifting the curve either in or out).

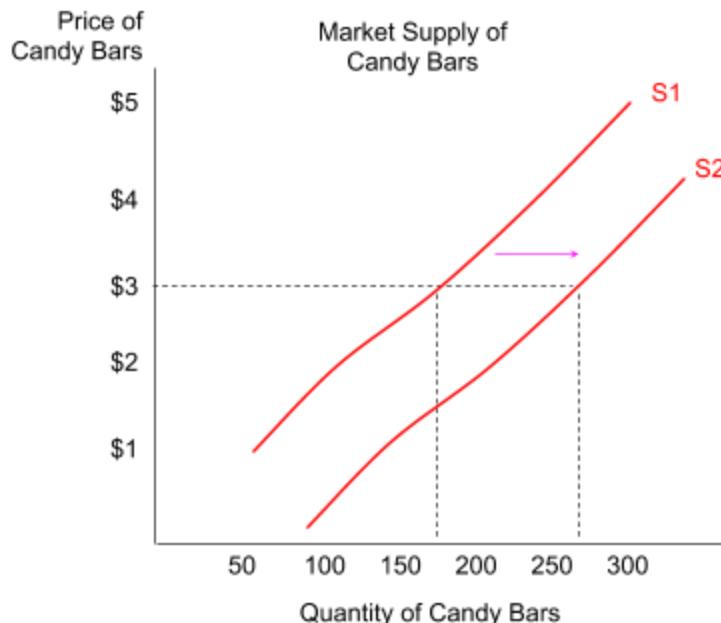
The determinants of supply

A change in price will lead to a change in the quantity demanded. But a change in a **determinant of supply** will shift the supply curve and cause more or less output to be supplied at EACH PRICE.

The non-Price Determinants of Supply (Supply shifters)	
Subsidies	Subsidies are government payments to producers for each unit produced. Subsidies cause supply to increase and the supply curve to shift outwards.
Taxes	Taxes are payments from producers to the government for each unit produced. Taxes cause supply to decrease and the supply curve to shift inward.
Technology	New technologies make production more efficient and increase supply.
Other related goods' prices	If another good that a firm could produce rises in price, firms will produce more of it and less of what they used to produce, decreasing supply. If substitutes in production get cheaper, firms will increase production of the now more profitable good, increasing supply
Resource costs	If the cost of inputs falls, supply will increase. If input costs rise, supply

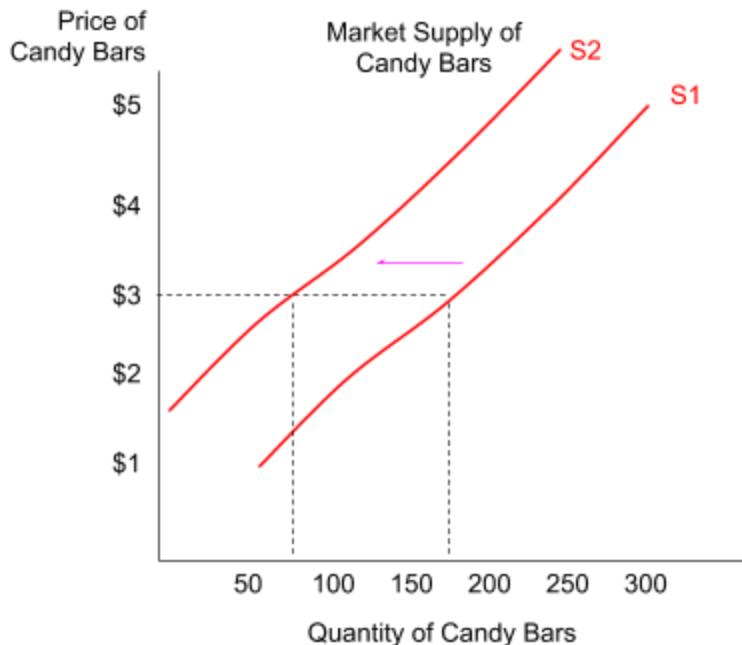
	decreases.
Expectations of producers	If firms expect the prices of their goods to rise, they will increase production now. If they expect prices to fall, they will reduce supply now.
Size of the market	If the number of firms in the market increases, supply increases. Vice versa.

Let's assume the price of chocolate, a main ingredient into candy bars, decreases. This will cause the supply of candy bars to increase, as seen in the graph below.



Due to lower input prices, suppliers are willing and able to provide more candy bars at every price. Now, instead of 160 candy bars at \$3, around 260 candy bars are supplied.

Let's look at the effect of a change in one more determinants of supply. Assume the government imposes a sugar tax on all candy sellers, forcing suppliers to pay the government \$1 in tax for each candy bar sold. This will shift the supply curve in and lead to a decrease in the supply.



Supply has decreased from S₁ to S₂. Now, at \$3 only 75 candy bars will be supplied, instead of 160. The tax has caused a decrease in supply; fewer candy bars are supplied at every price following the imposition of the tax.

Linear supply equations and graphs (HL only)

- Explain a supply function (equation) of the form $Q_s = c + dP$.
- Plot a supply curve from a linear function (eg, $Q_s = -30 + 20 P$).
- Identify the slope of the supply curve as the slope of the supply function $Q_s = c + dP$, that is d (the coefficient of P).
- Outline why, if the “c” term changes, there will be a shift of the supply curve.
- Outline how a change in “d” affects the steepness of the supply curve.

Linear Supply Equations

Supply can also be expressed mathematically as an equation. We will examine linear supply equations, which are simple formulas that tell us the quantity supplied of a good as a function of the good's price and non-price determinants.

A typical supply equation will be in the form:

$$Q_s = c + dP$$

Where:

- ‘ Q_s ’ = the quantity supplied for a particular good
- ‘ c ’ = the quantity supplied at a price of zero. This is the ‘q-intercept’ of supply, or where the supply curve would cross the Q-axis
- ‘ d ’ = the amount by which quantity will change as price changes, and
- ‘ P ’ = the price of the good

Consider the supply for bread in the same small village as in our demand analysis, which can be represented by the following equation:

$$Q_s = -200 + 150P$$

What do we know about the supply of bread from this function? We know that:

- If bread were free (e.g. if the price = 0), -200 loaves of bread would be demanded. Plug zero into the equation to prove that $Q_s = -200$ at a price of zero. Of course, -200 cannot be supplied, so if $P=0$, no bread will be produced.
- For every \$1 increase in the price of bread above zero, 150 additional loaves will be supplied. We can plug the following prices into the equation to prove this:

$$\begin{aligned} \$1 - Q_d &= -200 + 150(1) = -50 \\ \$2 - Q_d &= -200 + 150(2) = 100 \\ \$3 - Q_d &= -200 + 150(3) = 250 \\ \$4 - Q_d &= -200 + 150(4) = 400 \end{aligned}$$

We can also calculate the price at which the supply curve will begin. This is known as **the price-intercept** (because it's where the supply curve crosses the P-axis). To find this, set Q equal to zero and solve for P .

$$0 = -200 + 150(P)$$

$$P = 1.33$$

Creating a supply schedule from a supply equation

A supply equation can be plotted in both a supply schedule and as a supply curve. In the market for bread, we already determined the following:

- At a price of \$0, the quantity demanded is -200 loaves. This is the q-intercept
- At a price of \$1.33, the quantity supplied is 0 loaves. This is the p-intercept

With these numbers, we can create a supply schedule for the equation $Q_s = -200 + 150P$

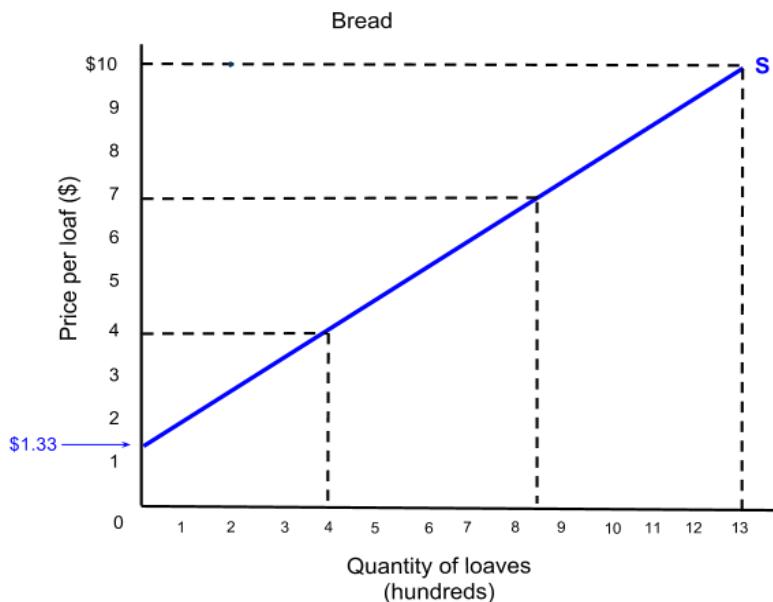
Price of bread	Quantity of loaves supplied
\$0	-200
\$2	100
\$4	400
\$6	700
\$8	1000
\$10	1300

Notice that as the price of bread rises from \$0 to \$10, the market goes from having no bread to having 1300 produced by firms.

For every \$1 increase in price, quantity supplied increases by 150 loaves; this corresponds with the 'd' variable, which is an indicator of the responsiveness of producers to price changes.

Plotting a supply curve from a supply schedule

The data from our supply schedule can easily be plotted on a graph. All we need is two points from the schedule to plot a curve. The following supply curve is for the equation $Q_s = -200 + 150P$



Notice the following:

- The Q-intercept is not visible on our graph, since the Q-axis only goes to the origin
- The P-intercept is labeled at \$1.33. This indicates that until the price of bread is \$1.33 per loaf, no firms will be willing to make bread.
- The steepness of the curve is affected by the 'd' variable, which tells us that for every \$1 increase in price, quantity rises by 150 loaves of bread. 'd' is the change in quantity over the change in price.

Changes in the 'c' variable

As we learned earlier, a change in price causes a change in the quantity supplied. This relationship can clearly be seen in the graph.

- But what could cause a shift in the supply curve?
- And how does this affect the supply equation?

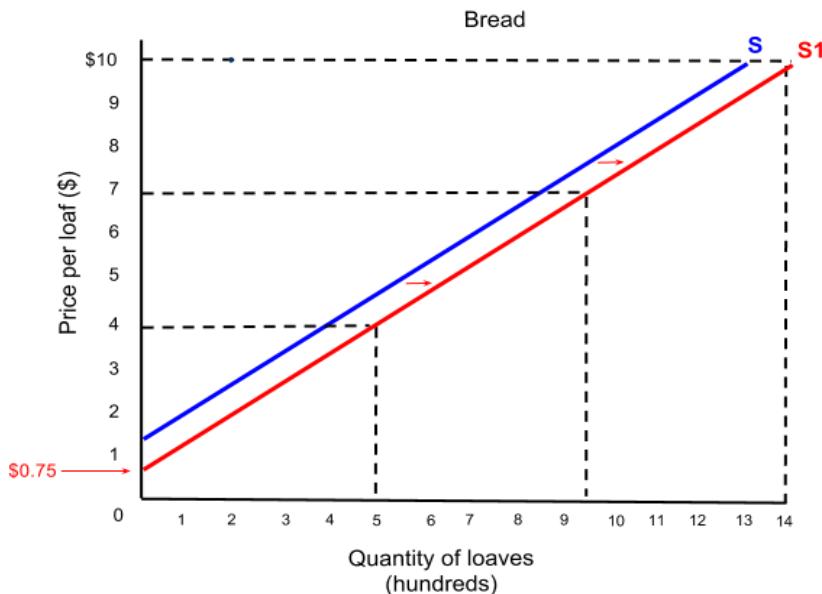
A change in a non-price determinant of supply will change the 'c' variable.

- Assume the price of wheat, a key ingredient in bread, falls.

- Supply of bread will increase and the supply curve will shift outward.
- In the supply equation, this causes the 'c' variable to increase. Assume the new equation is:

$$Q_s = -100 + 150P$$

Now more bread will be supplied at every price. The new Q-intercept is -100 loaves. The supply curve will shift to the right



Notice the following:

- At each price, 100 more loaves are now supplied. In the original graph, 400 loaves were supplied at \$4, now 500 are supplied.
- Supply has increased because a non-price determinant of supply changed (the price of an input decreased, so firms made more bread).
- The 'd' variable did not change, so the slope of the supply curve remained the same.
- The P-intercept decreased to \$0.75. Now, firms are willing to start baking bread at a price of just \$0.75, whereas before they would not begin making bread until the price reached \$1.33.

Changes in the 'd' variable

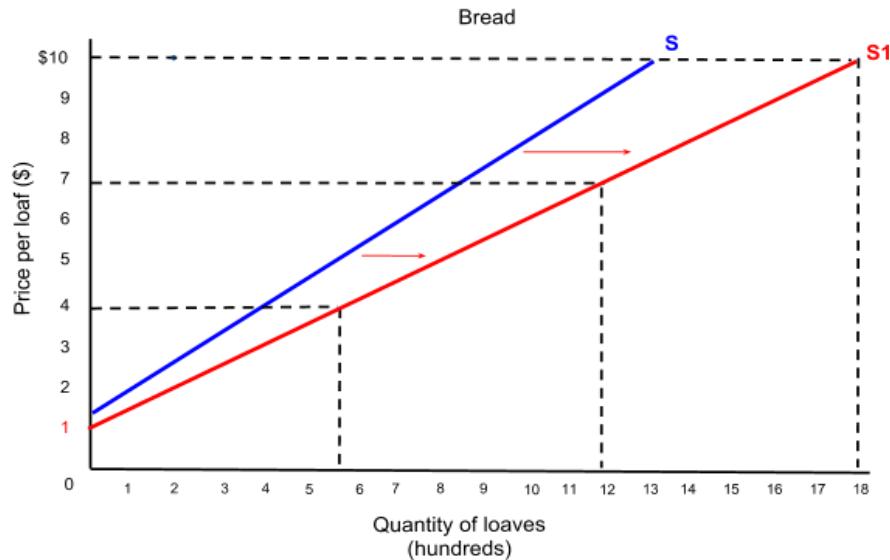
The 'd' variable in the supply equation is an indicator of the responsiveness of producers to price changes.

- If something causes producers to be more responsive to price changes, the 'd' variable will increase
- If something causes producers to be less responsive to price changes, the 'd' variable will decrease

Assume a new oven technology is developed that allows bakers to more quickly and efficiently increase their production of bread to satisfy rising demand from consumers. The 'd' variable in the supply equation increases as a result. The new equation is.

$$Q_s = -200 + 200P$$

Now, for every \$1 increase in price, producers will supply 200 more loaves, instead of 150. The Q-intercept will remain the same (-200) but the supply curve will be flatter, indicating producers are more responsive to price changes



Notice the following:

- Producers are more responsive to price changes now
- As the price rises from \$0 to \$4 per loaf, now producers will supply 600 loaves, whereas in the original graph they would have only supplied 400 loaves.
- Supply for bread has increased because bakers have acquired better technology.
- The new P-intercept at a lower price. It can be calculated by setting the Q to zero.

$$0 = -200 + 200P$$

$$P = 1$$

Now, at a price of \$1, firms will begin selling bread, whereas before the new oven technology, a price of \$1.33 was required.

Equilibrium and changes to equilibrium

- Explain, using diagrams, how demand and supply interact to produce market equilibrium.
- Analyse, using diagrams and with reference to excess demand or excess supply, how changes in the determinants of demand and/or supply result in a new market equilibrium.

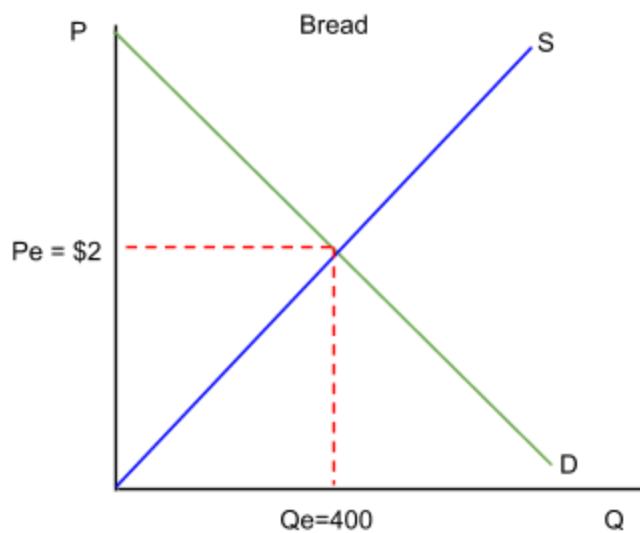
A market is in **equilibrium** when the price and quantity are at a level at which supply equals demand. The quantity that consumers demand is exactly equal to the quantity that producers

supply.

In equilibrium, a market creates no shortages or surpluses, rather, the market “clears”. Every unit of output that is produced is also consumed.

Equilibrium Price (Pe): The price of a good at which the quantity supplied is equal to the quantity demanded

Equilibrium Quantity (Qe): The quantity of output in at which supply equals demand.

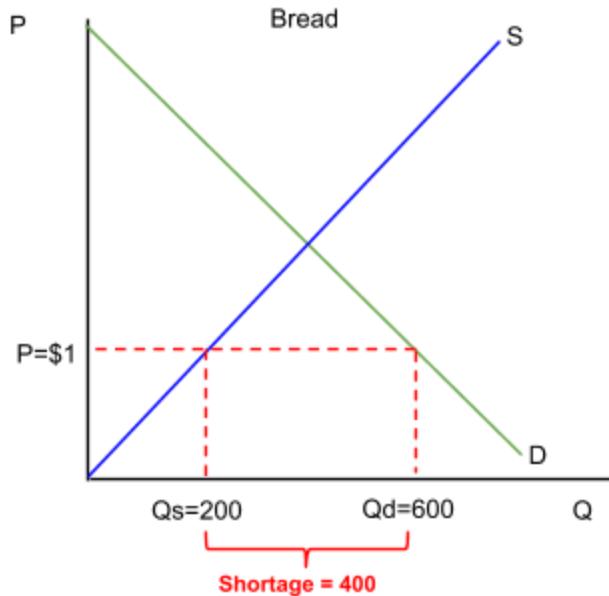


Consider the market for bread above:

- If the price were anything greater than \$2, firms would wish to supply more bread, but consumers would demand less. The market would be out of equilibrium.
- If the price were anything less than \$2, consumers would demand more but firms would make less. The market would be out of equilibrium.
- Only at \$2 does the quantity supplied equal the quantity demanded. This is the equilibrium point in the market for bread.

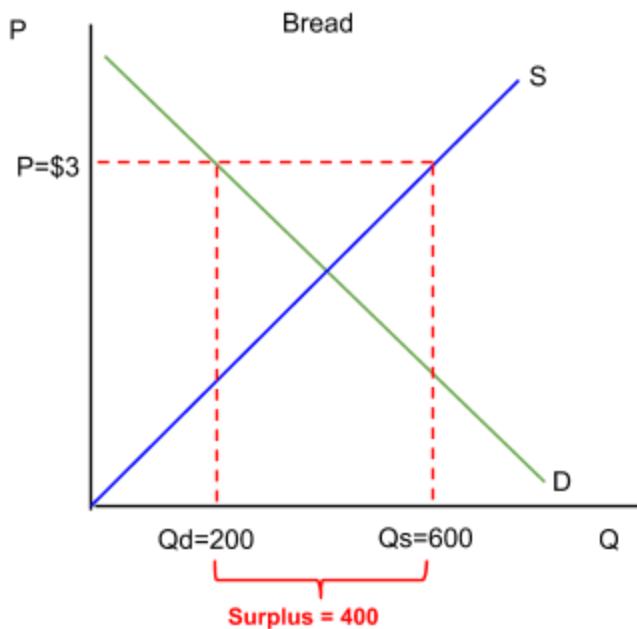
A market experiences a **disequilibrium** whenever the quantities demanded and supplied are different from one another. Disequilibrium results from a price that is either higher or lower than the equilibrium price.

For example, at \$1, the quantity demanded of bread is greater than the quantity supplied.



The market is in disequilibrium at \$1, as there is a **shortage** of 400 loaves of bread. This excess demand will go unmet at the price of \$1 per loaf.

When the price is above equilibrium, the quantity supplied will be greater than the quantity demanded.

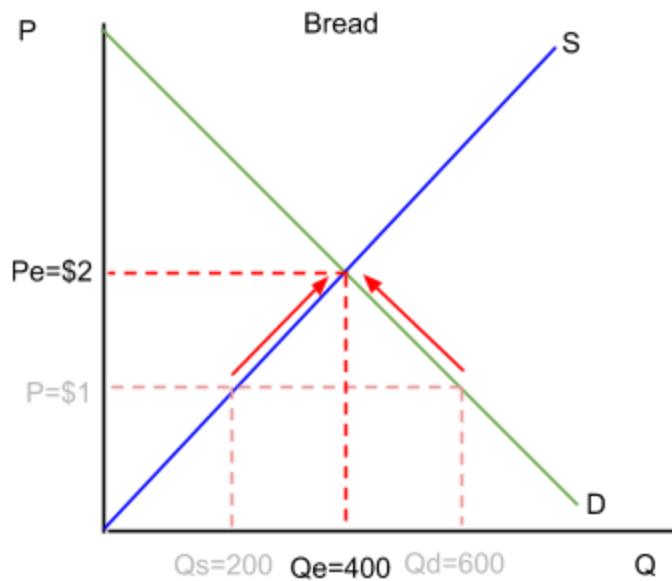


The market is in a disequilibrium at \$3, and there is a **surplus** of 400 loaves of bread. This excess supply will go unsold in the market at the price of \$4 per loaf.

Market forces will drive the price back to equilibrium when shortages and surpluses exist in a

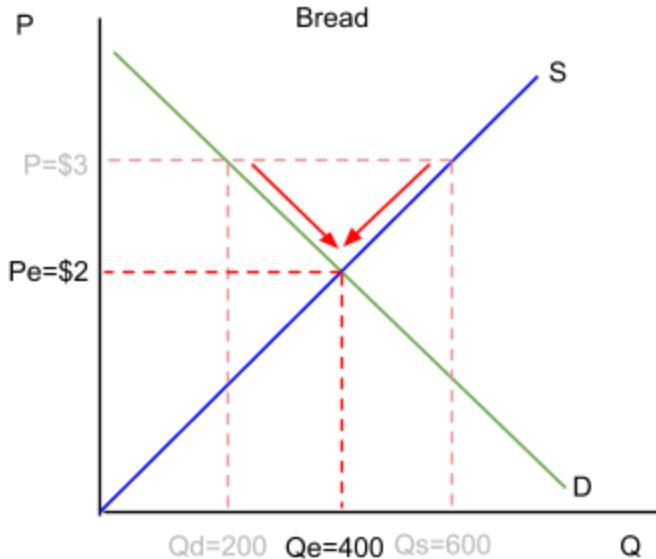
market.

For example, if the shortage created at the price of \$1 sends a signal to bakeries that consumers are willing to pay more for their bread. Sellers will thus raise their prices and increase the amount of bread they produce; at the same time, buyers will see the higher price and reduce the amount they demand until the quantities demanded and supplied are once again equal.



The shortage has driven the price back up to \$2, restoring equilibrium at 400 loaves of bread.

In the case of a surplus, excess supply will go unsold until sellers decide to lower their prices, at which time the quantity demanded will increase and the quantity supplied will decrease, restoring equilibrium.

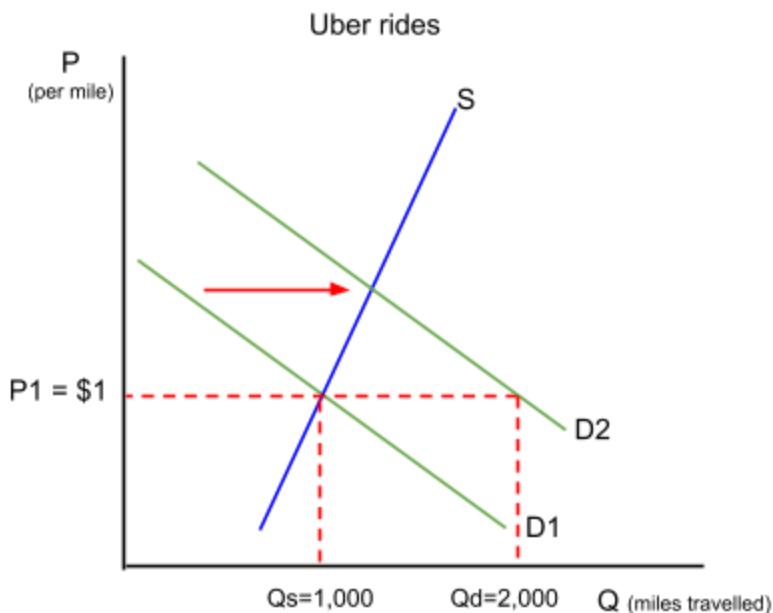


The surplus has caused the price to fall to \$2, restoring equilibrium at 400 loaves of bread.

Changes in market equilibrium

Factors that shift the market demand and market supply curves cause price, quantity, consumer surplus, producer surplus, and total economic surplus (within that market) to change. The impact of the change depends on the price elasticities of demand and supply.

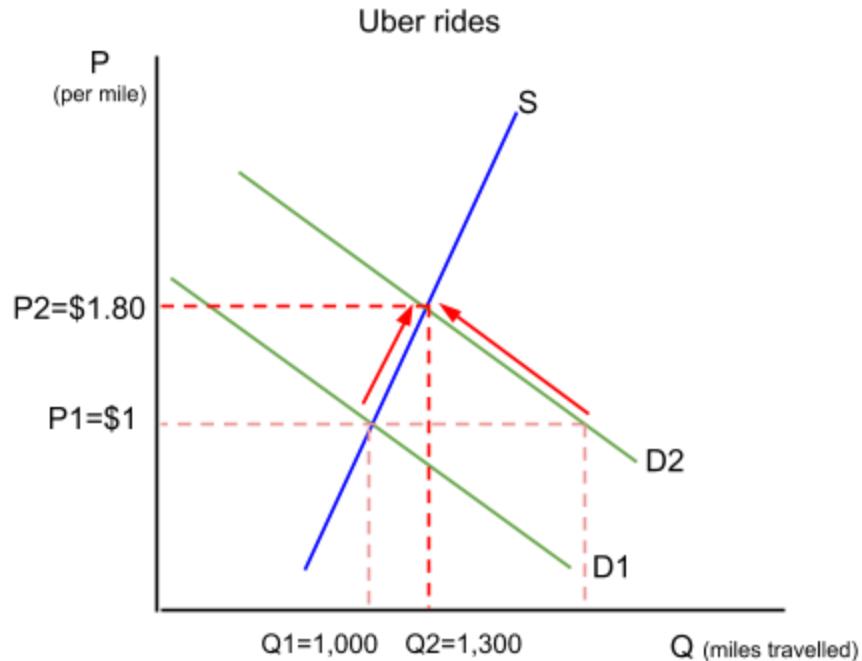
Consider the market for Uber rides in the graph below.



Demand for Uber rides has increased from D1 to D2, but the price has remained at \$1 per mile. There is a temporary shortage of Uber drivers. More people wish to ride an Uber than

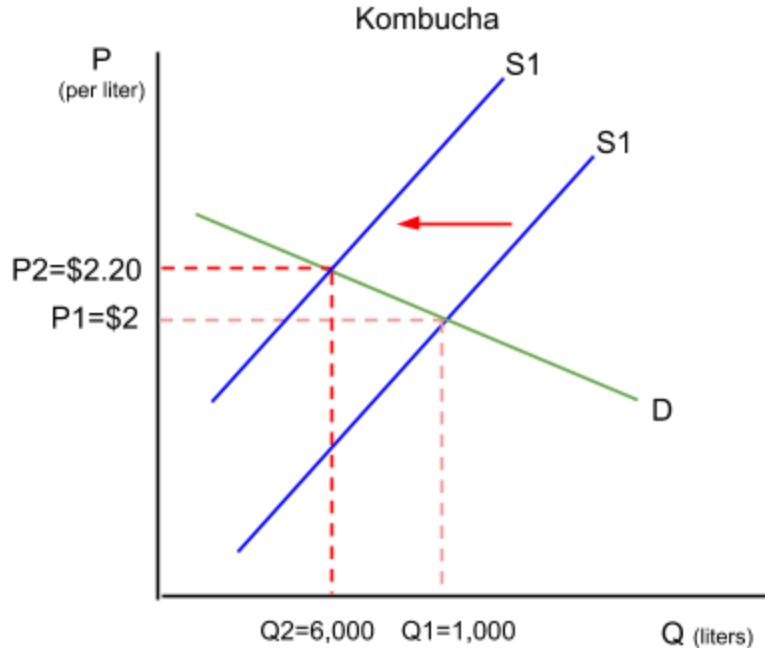
there are Uber rides available! How will the market adjust in response to this change in demand?

The shortage causes the price of Uber rides to rise, which in turn causes more drivers to offer rides, increasing the quantity supplied. At the same time, fewer people will wish to ride an Uber, and the quantity demanded will fall. Once these adjustments have taken place, the market will achieve a new equilibrium.



Notice that in this case, the supply of Uber rides is relatively inelastic. The result is the equilibrium quantity increases by relatively little compared to the equilibrium price. The 80% increase in price lead to only a 30% increase in quantity.

The impact of a change in demand or supply depends on the elasticities of demand and supply. Consider the market for Kombucha (a healthy beverage) below.



The supply of kombucha has decreased, causing the equilibrium price to rise and the quantity to fall. Since demand is relatively elastic, the change in quantity is proportionally larger than the change in price (40% decrease in quantity arising from a 10% increase in price).

Changes in demand and supply will cause the total surplus in a market to change.

- If demand or supply increase, there will be a greater quantity produced in a market and the total surplus among consumers and producers will increase.
- If demand or supply decrease, there will be a smaller quantity produced in a market and the total surplus among consumers and producers will decrease.
- The degree of the change in total surplus depends on the elasticities of demand and supply.
 - If supply and demand are relatively inelastic (consumers and producers are relatively unresponsive to price changes), then a change in demand or supply will have little impact on total surplus in the market.
 - If supply and demand are relatively elastic (consumers and producers are highly responsive to price changes), then a change in demand or supply will have a significant impact on total surplus in the market.

Calculating and illustrating equilibrium using linear equations (HL only)

- Calculate the equilibrium price and equilibrium quantity from linear demand and supply functions.
- Plot demand and supply curves from linear functions, and identify the equilibrium price and equilibrium quantity.
- Calculate the quantity of excess demand or excess supply in the above diagrams.

Equilibrium is a concept that can be transferred to our analysis of linear demand and supply

equations just as easily as it can be applied to graphs. Assume we have a market for bread in which demand and supply are represented by the equations:

$$Q_d = 600 - 50P \text{ and } Q_s = -200 + 150P$$

Equilibrium price and quantity occur when demand equals supply. So to calculate the equilibrium using these equations, we must set the two equal to each other and solve for price

$$600 - 50P = -200 + 150P$$

$$800 = 200P$$

$$P = \$4$$

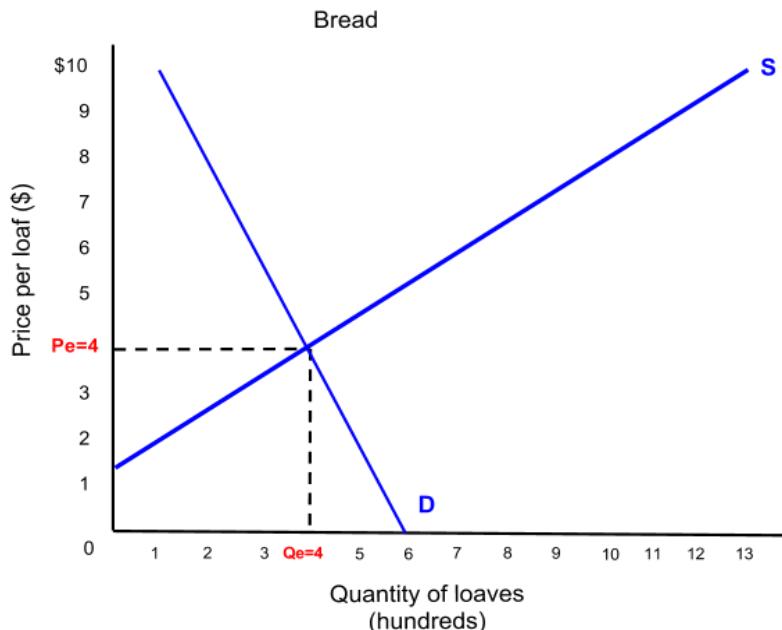
Next, to find the equilibrium quantity, we must simply put the \$4 price into either the demand or supply equation (since they will both yield the same quantity)

$$Q_d = 600 - 50(4)$$

$$Q_d = 400$$

The equilibrium price of bread is **\$4** and the equilibrium quantity is **400 loaves**

If we plot the demand and supply curves on the same axis, the intersection of the two curves should confirm our calculations of equilibrium price and quantity.



Notice:

- If the price were anything other than \$4, the quantities demanded and supplied

- would not be equal.
- If the quantity were anything other than 400, the marginal social benefit (demand) and marginal social cost (supply) would not be equal.
 - \$4 is the market clearing price and 400 is the efficient level of output.

Changes to market equilibrium

Assume the cost of producing bread rises (perhaps wages for bakers have increased). The supply of bread will decrease and the supply equation changes to:

$$Q_s = -400 + 150P$$

Assume demand remains at $Q_d = 600 - 50P$. What will the decrease in supply do to the market equilibrium price and quantity? We can calculate the new equilibrium easily:

$$600 - 50P = -400 + 150P$$

$$1000 = 200P$$

$$P = \$5$$

The decrease in supply made bread scarcer and caused the price to rise. The quantity should decrease, which we can confirm by solving for Q.

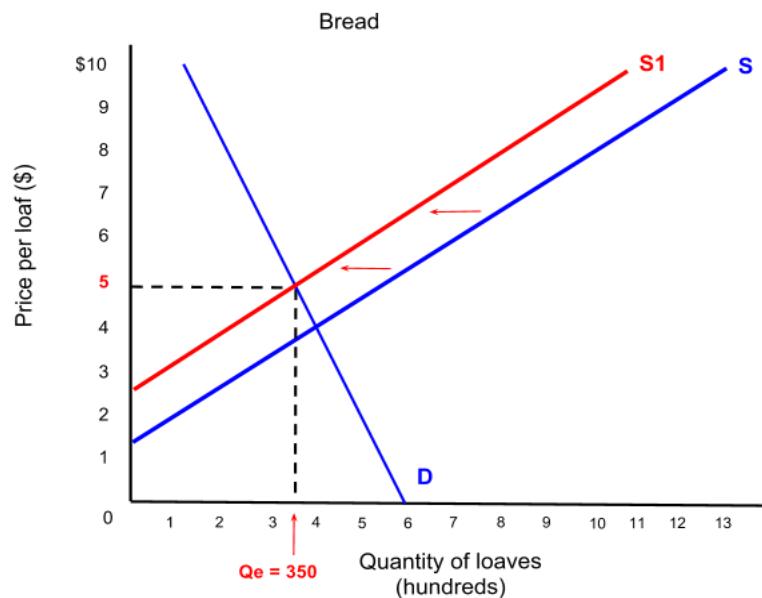
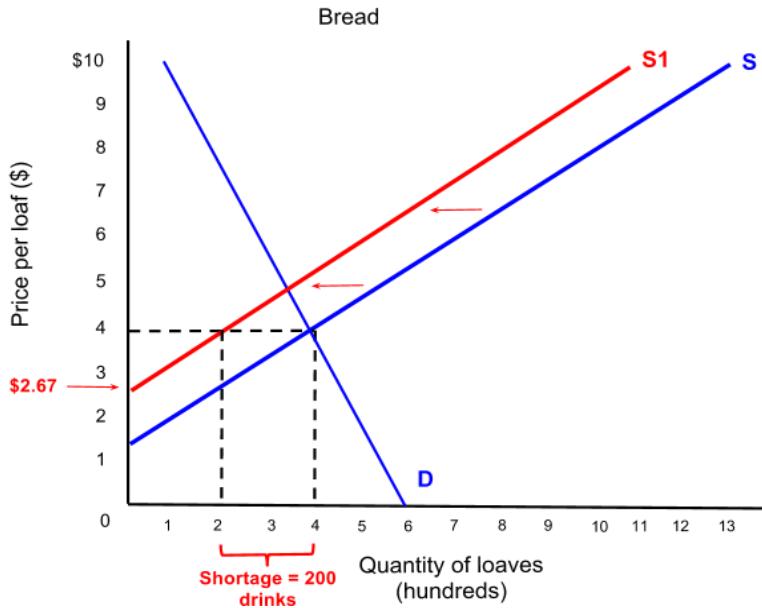
$$Q_d = 600 - 50(5)$$

$$Q_d = 350$$

A decrease in supply caused the equilibrium price to rise and the quantity to decrease in the market for bread!

As the supply decreases, the price of bread must rise, or else there will be shortages (as seen in the first graph). Once the market adjusts to its new equilibrium, the shortages are eliminated and the Q_d once again equals the Q_s (as seen in the second graph).

$$Q_s = -400 + 150P \text{ and } Q_d = 600 - 50P$$

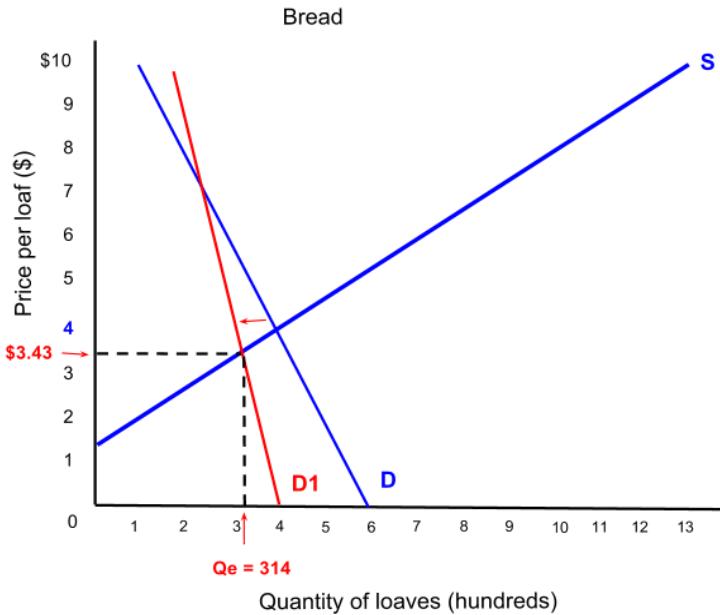


What if the demand changes? Assume consumers become less responsive to change in the price of bread and the demand equation changes to:

$$Q_d = 400 - 25P$$

while supply remains the same at $Q_s = -200 + 150P$

If we graph these two equations, we can see the new equilibrium price and quantity



Notice:

- Demand has decreased and become steeper, indicating that consumers are less responsive to price changes, yet consumer a smaller quantity overall.
- The equilibrium price is lower (\$3.43 instead of \$4) and the quantity is lower (314 instead of 400)

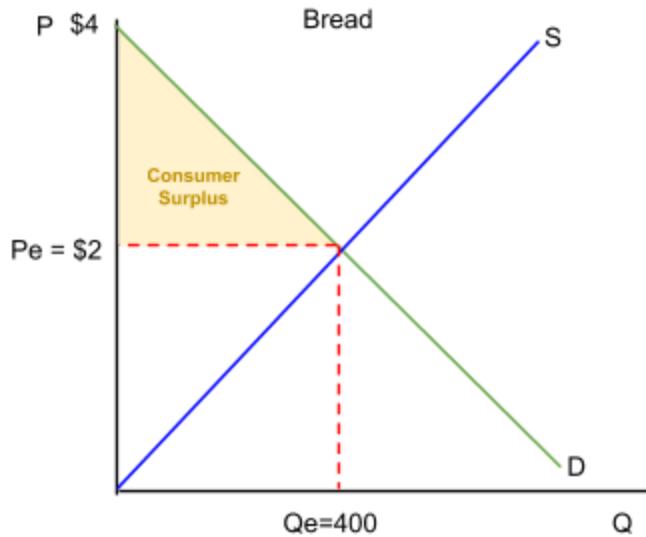
Whenever either demand or supply change, the market equilibrium will adjust to a new market clearing price and quantity!

Consumer surplus

- Explain the concept of consumer surplus.
- Identify consumer surplus on a demand and supply diagram.

Consumer surplus refers to the benefit enjoyed by consumers who were willing to pay a higher price than they had to for a good. The demand curve represents the willingness of consumers to pay for a good. Any consumer willing to pay more than the equilibrium price enjoys “surplus happiness”, or extra happiness, because they feel like they got a good deal!

In the market for bread, the total consumer surplus is represented by the area below the demand curve and above the equilibrium price.



Consumer surplus can be quantified in dollars by finding the area of the yellow triangle:

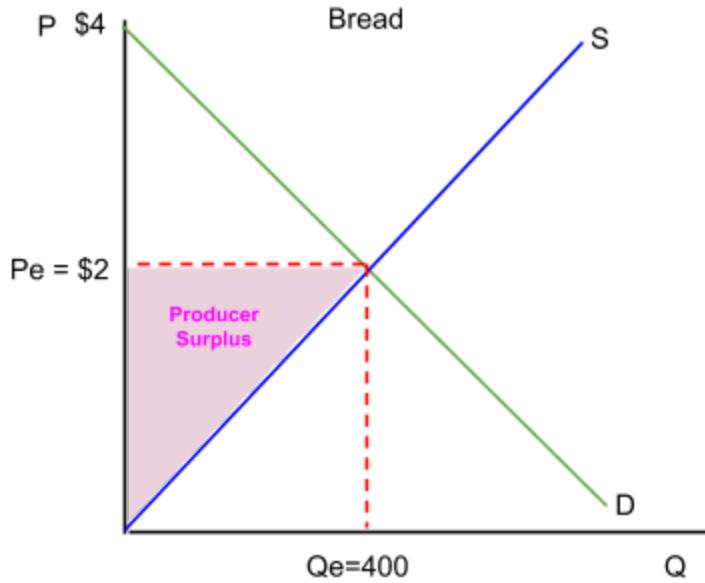
- Consumer surplus = $((\$4 - \$2) \times 400) \div 2 = \400
- The value of the consumer surplus enjoyed by bread buyers who paid less than they were willing to pay is \$400.

Producer surplus

- Explain the concept of producer surplus.
- Identify producer surplus on a demand and supply diagram.

Producer Surplus is the benefit enjoyed by producers who would have been willing to sell their product at a lower price than they were able to. Because some producers are able to produce a good at a cost lower than the price they sell it for, they enjoy “surplus happiness” by selling it at a price that is profitable to them.

In the market for bread, the producer surplus is represented by the area below the equilibrium price and above the supply curve.



The value of producer surplus can be calculated by finding the area of the triangle of the pink triangle:

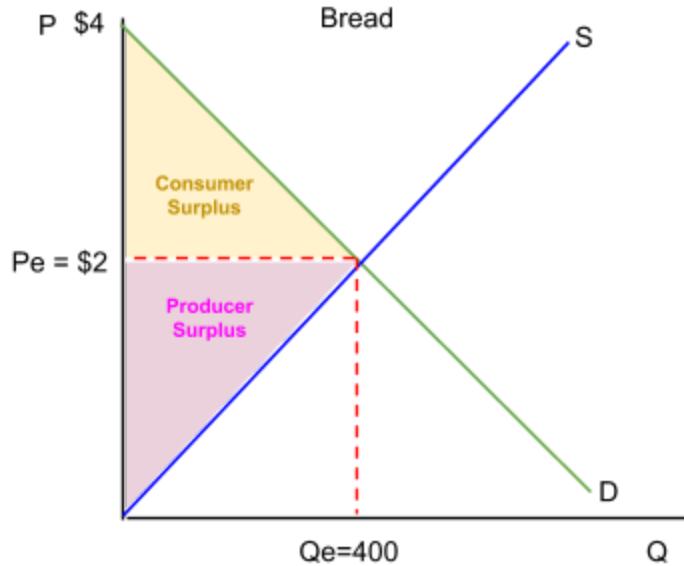
- Producer surplus = $((\$2-\$0) \times 400) \div 2 = \400
- The value of the producer surplus enjoyed by bakers who are able to sell bread at a price higher than they would have been willing to sell it is \$400.

Allocative Efficiency

- Evaluate the view that the best allocation of resources from society's point of view is at competitive market equilibrium, where social (community) surplus (consumer surplus and producer surplus) is maximized (marginal benefit = marginal cost).

The **total surplus**, or welfare, in a market is the sum of producer and consumer surplus. Total surplus represents the benefits of all the agents in a market who feel like they're getting a good deal at the equilibrium price, including the buyers who are paying less than they'd have been willing to pay and the sellers who sell for more than they'd have been willing to sell for.

Graphically, total surplus is the area below the demand curve and above the price, and above the supply curve and below the price.



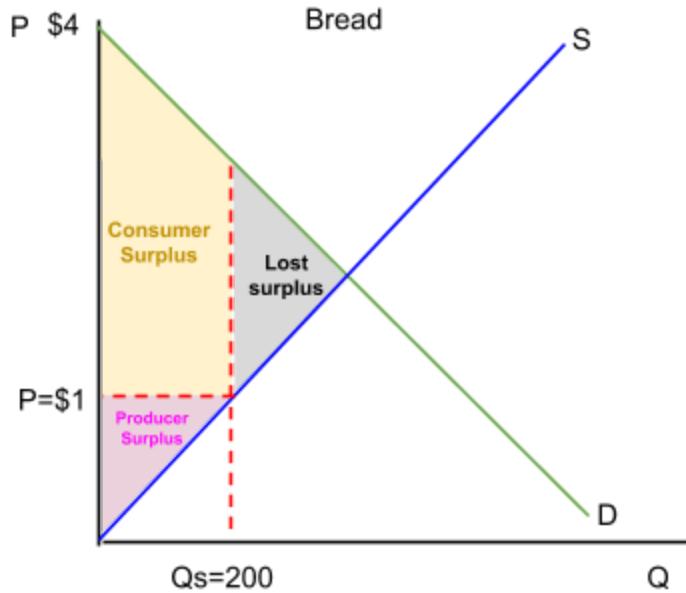
Total surplus can be calculated by adding the values of consumer and producer surplus.

- Total surplus = $\$400 + \$400 = \$800$
- The value of the benefits enjoyed by all the producers and consumers of bread who feel like they're getting a good deal at a price of \$2 is \$400.

Total surplus, market equilibrium, and allocative efficiency

A competitive market in equilibrium achieves the maximum amount of total surplus possible, assuming there are no external costs or benefits arising from the good's production or consumption.

At any other price/quantity combination, the total value of consumer and producer surplus will be less than that at equilibrium. For example, assume bread prices were decreased to \$1, without any change in demand or supply.



At any price other than equilibrium, total surplus is reduced and there is a **deadweight loss** in the market.

A market in equilibrium is **allocatively efficient** because it maximizes consumer and producer surplus.

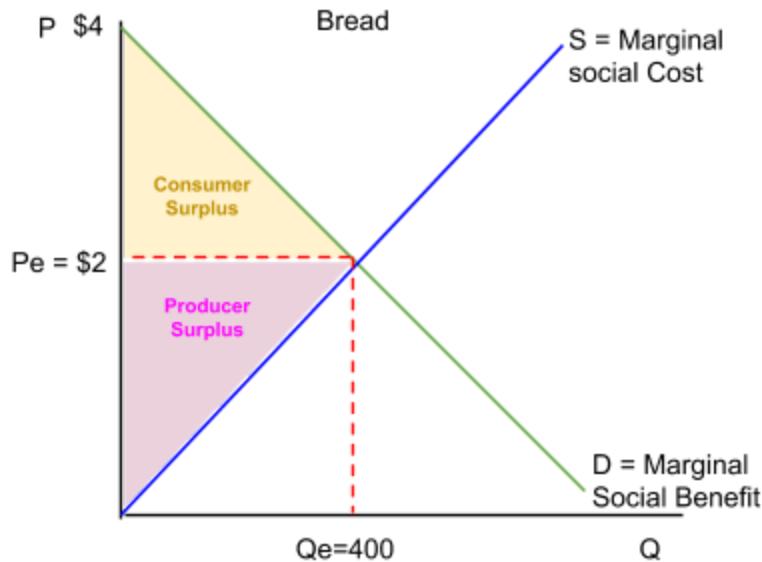
- Neither too much nor too little is being produced.
- Every unit that is produced is consumed, and the market clears.

When a market is in equilibrium, resources are efficiently allocated. To understand what this means, we must think about demand and supply in a new way.

- Demand = Marginal Social Benefit (MSB): The demand for any good represents the benefits that society derives from the consumption of that good. Marginal benefits decrease at higher levels of output because additional units of a good bring benefits to fewer and fewer people the more of the good exists.
- Supply = Marginal Social Cost (MSC): The supply of a good represents the cost to society of producing the good. For almost all goods, the greater the amount is produced, the more it costs to produce additional units of it. Think of oil. As the world produces more and more oil, it becomes increasingly harder to produce, thus the marginal cost (the cost for each additional barrel) continuously rises.

Only when the MSB = MSC is society producing the right amount of any good. If output occurs at any other level, we must say that resources are misallocated towards the good.

Once again, consider the market for bread below.



The value society places on the 400th loaf (the MSB) is equal to the cost of society producing it (the MSC).

- At any level of output below 400 loaves of bread, society's benefit (MSB) would be greater than society's cost (MSC), and MORE should be produced.
- At any level of output above 400 loaves of bread, society's cost (MSC) would be greater than society's benefit (MSB), and LESS should be produced.
- Only at equilibrium (where supply equals demand) are resources efficiently allocated; e.g. neither too much nor too little is produced. The market is efficient because the marginal benefit of the last unit produced equals the marginal cost of the last unit produced.

Allocative efficiency is achieved in a market when the quantity is produced at which the benefit society derives from the last unit is equal to the cost imposed on society to produce the last unit.

Allocative efficiency is achieved where Marginal Social Benefit = Marginal Social Cost

Assuming there are no “external” costs or benefits from the production or consumption of a good, a free market will achieve allocative efficiency when the equilibrium price and quantity prevail.

1.2 Elasticities

Price elasticity of demand and its determinants

- Explain the concept of price elasticity of demand, understanding that it involves responsiveness of quantity demanded to a change in price, along a given demand curve.
- Calculate PED using the PED equation.
- State that the PED value is treated as if it were positive although its mathematical value is usually negative.
- Explain, using diagrams and PED values, the concepts of price elastic demand, price inelastic demand, unit elastic demand, perfectly elastic demand and perfectly inelastic demand.
- Explain the determinants of PED, including the number and closeness of substitutes, the degree of necessity, time and the proportion of income spent on the good.
- Calculate PED between two designated points on a demand curve using the PED equation.
- Explain why PED varies along a straight line demand curve and is not represented by the slope of the demand curve

Introduction to elasticities

Elasticity is an economic concept that refers to the responsiveness among consumers or producers to a change in a variable that affects either the market demand or the market supply. There are four types of elasticity that we will study in this unit:

- **Price Elasticity of Demand (PED):** Measures the responsiveness of consumers of a particular good to a change in the good's price.
- **Price elasticity of Supply (PES):** Measures the responsiveness of producers of a particular good to a change in the price of that good.
- **Cross-price elasticity of Demand (XED):** Measures the responsiveness of consumers of one good to a change in the price of a related good (either a substitute or a complement).
- **Income Elasticity of Demand (YED):** Measures the responsiveness of consumers of a particular good to a change in their income.

Price elasticity of demand – definition and formula

Price Elasticity of Demand (PED) is a measurement of how much the quantity demanded for a good will change as a result of a particular change in the good's price. PED can range from a value of 0 to infinity, and is calculated using the following formula:

$$\text{PED} = \frac{\text{The percentage change in the quantity of a good demanded}}{\text{The percentage change in the price of the good, or}}$$

$$\text{PED} = \frac{\% \Delta Q_d}{\% \Delta P}$$

If, for example, we know that an increase in the price of bananas from \$4 to \$6 caused the quantity demanded to fall from 1,000 bananas to 800 bananas, we can calculate the PED for bananas.

$$\% \Delta Q_d = (800 - 1000) \div 1000 = -0.2 \times 100 = -20\%$$

$$\% \Delta P = (6 - 4) \div 4 = 0.5 \times 100 = 50\%$$

$$PED = (-20) \div 50 = -0.4$$

Notice that our PED has a negative value.

- This reflects the law of demand
- Whichever direction the price of good moves in, the quantity will always move in the opposite direction
- Since PED will always be negative, we can refer to it in its absolute value. So, the PED for bananas is 0.4.

Interpretation of the PED coefficient:	
If PED is less than 1:	Demand is inelastic . This means that the percentage change in the quantity is less than the percentage change in the price.
If PED is greater than 1	Demand is elastic . The percentage change in the quantity is greater than the percentage change in the price.
If PED=0:	Demand is perfectly inelastic . There was no change in quantity resulting from the price change.
If PED=1:	Demand is unit elastic . The percentage change in the quantity was identical to the percentage change in the price.
If PED = infinity:	Demand is perfectly elastic . The smallest increase in price causes the quantity demanded to fall to ZERO.

Interpretation of PED

Demand for bananas was 0.4. Based on our interpretations of PED from the table above, we know that demand for bananas is inelastic.

- For every 1% increase in the price of bananas between \$4 and \$6, the quantity demanded fell by 0.4%.
- Since price increased by a total of 50%, the quantity fell by a total of just 20%.
- Consumers are relatively unresponsive to the price of bananas.

The determinants of PED

Whether demand for a good at a particular price is elastic or inelastic depends on several characteristics of the good itself. Just how much will consumers respond to a price change for the good? The following table presents some of the primary **determinants of PED**.

S	Substitutes	The number of substitutes available. The more substitutes, more elastic demand, as consumers can replace a good whose price has gone up with one of its now relatively cheaper substitutes.
P	Proportion of income	The proportion of income the purchase of a good represents. If a good represent a higher proportion of a consumer's income, his demand tends to be more elastic.
L	Luxury or necessity?	Luxury or necessity? If a good is a necessity, changes in price tend not to affect quantity demand, i.e. demand is inelastic. If it's a luxury that a consumer can go without, consumers tend to be more responsive.
A	Addictive?	If a product is addictive or habit forming, demand tends to be inelastic.
T	Time	The amount of time a consumer has to respond to the price change. If prices remain high over a longer period of time, consumers can find substitutes or learn to live without, so demand is more elastic over time.

Applications of price elasticity of demand

- Examine the role of PED for firms in making decisions regarding price changes and their effect on total revenue.
- Explain why the PED for many primary commodities is relatively low and the PED for manufactured products is relatively high.
- Examine the significance of PED for government in relation to indirect taxes.

The PED formula is useful for more than just telling us how much consumers respond to price changes. It can be very useful to businesses and government decision-making.

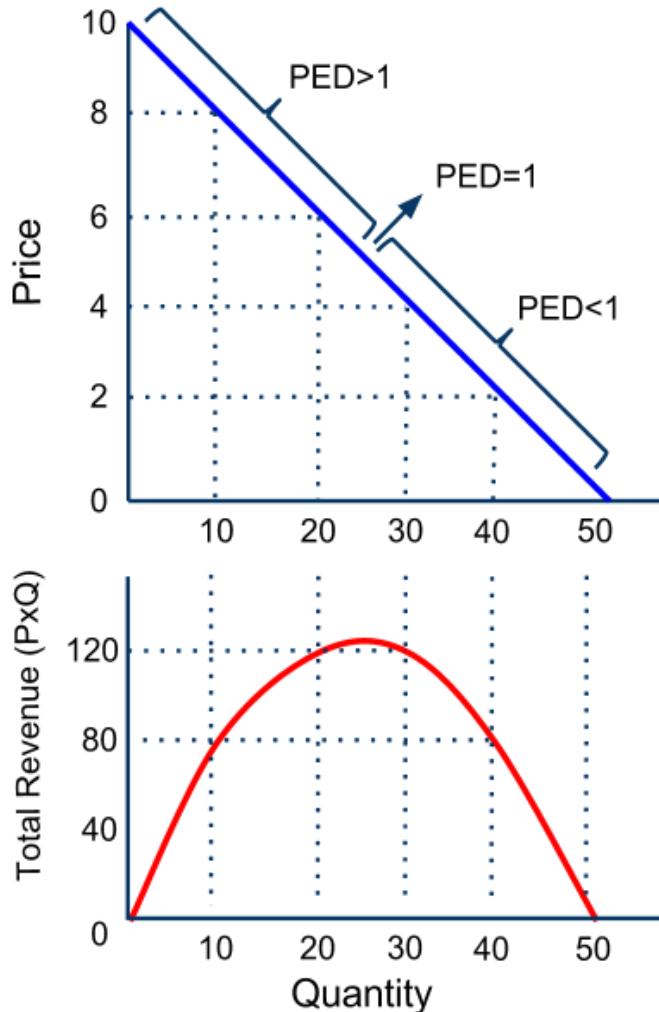
Applications of PED for	
Businesses	Businesses benefit from knowing how responsive their consumers are to price changes at any given time. <ul style="list-style-type: none"> • If a seller knows demand is HIGHLY elastic, he may wish to lower the price and capture many new customers. • If a seller knows demand is highly inelastic, he may wish to raise his price as he will not lose many sellers but will enjoy higher revenues.
Government	The government needs to know how consumers will respond to taxes imposed on particular goods. For example, if the government wishes to raise revenues from taxing goods, it should know that: <ul style="list-style-type: none"> • A tax on restaurant meals (relatively elastic) will not raise much revenue because people will just stop going to restaurants. • A tax on cigarettes (relatively inelastic) will raise lots of revenue because most people will continue smoking and thus have to pay the tax.

The total revenue test of PED

A quick way to determine whether demand is elastic or inelastic between two prices is to determine whether the sellers revenues (or consumers expenditures) rise or fall as a result of a price change. Total revenue (or expenditures) is calculated as the price times the quantity

$(P \times Q)$.

The graphs below show a good's demand and the total revenues of its sellers at each of the prices from \$10 to \$0.



We can calculate total revenue at each of the prices:

- At \$10: $TR=10 \times 0=0$
- At \$8: $TR=8 \times 10=80$
- At \$6: $TR=6 \times 20=120$
- At \$4: $TR=4 \times 30=120$
- At \$2: $TR=2 \times 40=80$
- At \$0: $TR=0 \times 50=0$

By looking at how a change in price causes total revenues to change, we can determine whether demand is inelastic or elastic between two prices.

- If a decrease in price causes TR to rise, demand is elastic.
- If a decrease in price causes TR to fall, demand is inelastic
- If an increase in price causes TR to rise, demand is inelastic

- If an increase in price causes TR to fall, demand is elastic

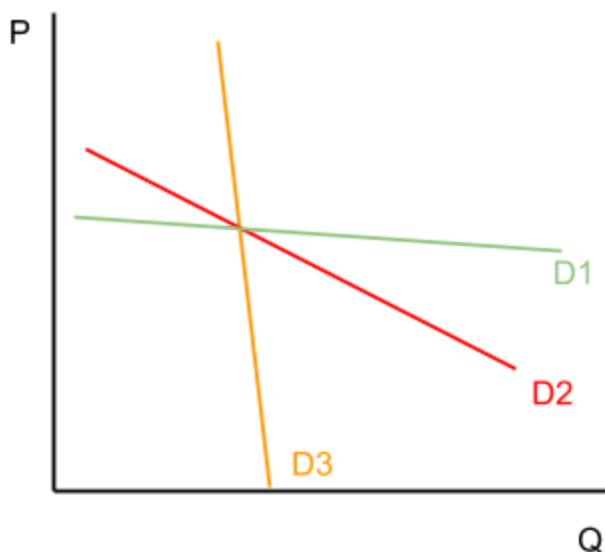
Based on the test above, we can see that demand is elastic as price falls from \$10 to \$5, unit elastic at \$5, and inelastic below \$5. Consumers are relatively responsive to price changes when price is high and quantity is low; but they are less responsive to price changes when price is low and quantity is high.

The **total revenue test** is helpful for businesses considering changing their prices. For instance, if a fast food restaurant wants to raise its prices, but knows that demand for its food is elastic, then the restaurant could be sure that a higher price will cause its revenues to decrease. On the other hand if they know demand is elastic, then a price cut would cause their revenues to increase.

PED and the slope of the demand curve

PED and slope are different concepts.

- Slope of a line measures the rise over the run, or in the demand curve the change in price over the change in quantity.
- PED measures the percentage change in quantity over the percentage change in price.
- However, by comparing the relative slopes of demand curves plotted on the same axis, we can determine the relative elasticity of different goods.



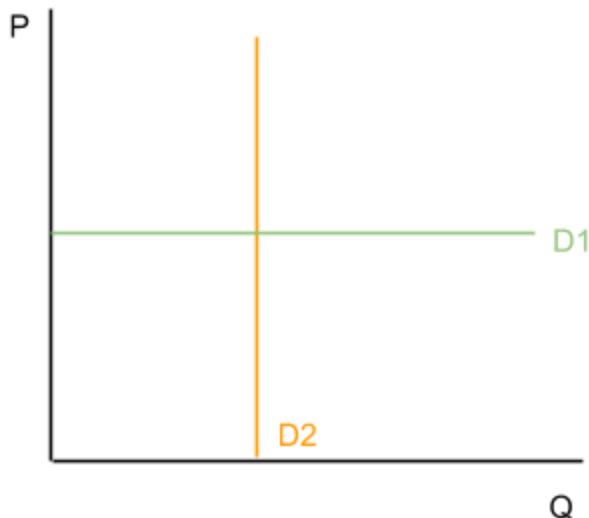
Between any two price, consumers of the good represented by D1 will be more responsive than consumers of the goods represented by D2 and D3. The steeper the demand curve on a particular axis, the more elastic demand for that good is between any two prices.

Perfectly elastic and perfectly inelastic demand

In some rare cases, any increase in price, no matter how small, will result in the quantity of a good demanded falling to zero. In such a case demand is perfectly elastic.

On the other hand, if the quantity demanded of a good does not change no matter how much price increases, demand is perfectly inelastic.

- If the value of the elasticity coefficient = ∞ , demand is perfectly elastic and the demand curve is horizontal.
- If the value of the elasticity coefficient = 0, demand is perfectly inelastic and the demand curve is vertical.



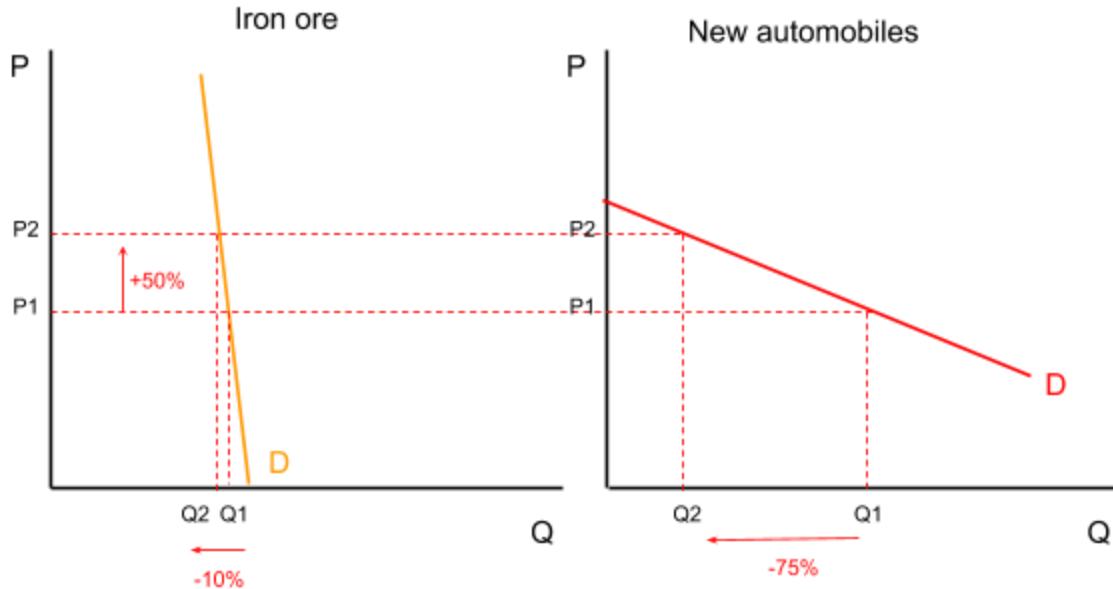
D1 in the graph above is perfectly elastic. Any change in price will cause quantity to change infinitely in the opposite direction. D2 is perfectly inelastic. Any change in price will have no impact on the quantity demanded.

PED and primary vs. manufactured goods

Whether a good is a primary commodity or a manufactured good is likely to impact whether its demand is relatively elastic or inelastic.

- Primary commodities, such as energy and mineral resources, agricultural commodities, and other raw materials used in the production of goods tend to have relatively inelastic demand. They have few substitutes and their prices often make up a small proportion of the value of the finished goods they are used to produce.
- Manufactured goods tend to have relatively more substitutes than the primary goods that go into their production, and therefore demand for manufactured goods tends to be relatively inelastic.

The graphs below show the demand for iron ore, a primary commodity, and new cars, a manufactured good for which iron ore is included in the production.



Notice:

- The demand for iron ore is relatively inelastic: $PED = \frac{-10\%}{50\%} = -0.2$
- The demand for new cars is relatively elastic: $PED = \frac{-75\%}{50\%} = -1.5$

New cars have a nearly perfect substitute: used cars. Therefore, when new car prices increase by 50%, buyers in large numbers switch to used cars.

Iron ore, on the other hand, has no substitutes when it comes to its primary use as an industrial input. Therefore, a 50% increase in iron prices only causes a 10% fall in quantity demanded.

Demand for primary commodities is relatively inelastic compared to their manufactured counterparts.

PED and a government's taxation of goods and services

Governments interested in raising revenues to finance government spending often decide to tax consumption of particular goods and services. Taxing goods for which demand is relatively inelastic will generate more revenue than taxing goods that have a relatively elastic demand.

- A tax on an inelastic good will have relatively little impact on the quantity of the good consumed. Therefore, tax revenues will be relatively large.
- A tax on an elastic good will lead to a proportionally large fall in the quantity of the good consumed. Therefore, rather than raising significant amounts of revenue, taxing elastic goods will just wipe out demand for that good and reduce the amount of both consumer and producer surplus.

A graphical analysis of the importance of PED in regards to consumption taxes will be explored in a later unit.

Cross price elasticity of demand and its determinants

- Explain the concept of cross price elasticity of demand, understanding that it involves responsiveness of demand for one good (and hence a shifting demand curve) to a change in the price of another good.
- Calculate XED using the XED equation.
- Show that substitute goods have a positive value of XED and complementary goods have a negative value of XED.
- Explain that the (absolute) value of XED depends on the closeness of the relationship between two goods.
- Examine the implications of XED for businesses if prices of substitutes or complements change.

Cross price elasticity of demand (XED) measures the responsiveness of consumers of one good to a change in the price of a related good.

For example, consider apples and pears, two fruits that are close substitutes for one another.

- How will demand for pears be affected by an increase in the price of apples?
- XED tells us the percentage by which quantity of pears demanded will change following a particular percentage change in the price of apples.

XED = The percentage change in the quantity of one good ÷ The percentage change in the price of a related good, or

$$\text{XED} = \frac{\% \Delta Q_d \text{ of good A}}{\% \Delta P \text{ of good B}}$$

Assume the following:

- The price of apples rises from \$2 to \$2.50
- The quantity of pears demanded rises from 30 to 50

$$\begin{aligned}\text{XED of apples and pears} &= ((50-30) \div 30) \div ((2.5-2) \div 2) \\ &= 0.67 \div 0.25 = 2.7\end{aligned}$$

Demand for pears is cross price elastic with apples (i.e. XED>1)

Just like PED, the absolute value of XED can be:

- 0-1: Inelastic – Consumers of Good A are relatively unresponsive to a change in the price of Good B (the % change in Q_A will be smaller than the % change in P_B)
- 1: Unit Elastic – Consumers of Good A will respond proportionally to a change in the price of Good B (the % change in Q_A will be the same as the % change in P_B)
- >1: Elastic – Consumers of Good A will be relatively responsive to a change in the price of Good B (the % change in Q_A will be greater than the % change in P_B)

XED can be either negative or positive

The XED for complementary goods will always be negative, because when the price of one

complement goes up, the demand for the other will fall.

- Example: When the price of hot dogs rises, the demand for hot dog buns will decrease. XED coefficient will be negative

The XED for substitutes will always be positive, because when the price of one substitute goes up, the demand for the other will rise.

- Example: When the price of beef rises, the demand for pork will rise. XED coefficient will be positive, reflecting the direct relationship

Income elasticity of demand and its determinants

- Explain the concept of income elasticity of demand, understanding that it involves responsiveness of demand (and hence a shifting demand curve) to a change in income.
- Calculate YED using the YED equation.
- Show that normal goods have a positive value of YED and inferior goods have a negative value of YED.
- Distinguish, with reference to YED, between necessity (income inelastic) goods and luxury (income elastic) goods.
- Examine the implications for producers and for the economy of a relatively low YED for primary products, a relatively higher YED for manufactured products and an even higher YED for services.

Income elasticity of demand (YED) measures the responsiveness of consumers of a good to a change in the level of their income.

For example, imagine a country is going into recession, so the income of the average household is falling. Demand for new cars is falling, but demand for bicycles is rising. YED is a measure of how responsive consumers' demand for bicycles and cars is to changes in their incomes.

YED = Percentage change in the quantity of a good ÷ Percentage change in consumer's income, or

$$\text{YED} = \frac{\% \Delta Q_d}{\% \Delta Y}$$

Assume the following:

- Incomes in America have fallen by 4%
- Bike sales have risen by 8%
- Car sales have fallen by 3%

$$\text{YED for bikes} = 8 \div -4 = -2$$

Demand for bikes is income elastic

$$\text{YED for cars} = -3 \div -4 = 0.75$$

Demand for cars is income inelastic

As with PED and XED, the absolute value of YED can be:

- 0-1: Inelastic – Demand for the good is relatively unresponsive to changes in consumer income (quantity will change by a smaller percentage than the change in income)
- 1: Unit Elastic – Demand for the good is proportionally responsive to income changes (quantity will change by the same percentage as the change in income)
- >1: Elastic – Demand for the good is relatively responsive to changes in income (quantity will change by a larger percentage than consumers' income)

YED can be either positive or negative

A **normal good** is one with a positive YED coefficient. There is a direct relationship between income and demand.

- Example: As incomes fell, car sales fell as well. If incomes were to rise, car sales would begin to rise. Cars are a normal good, so the YED coefficient is positive.

An **inferior good** is one with a negative YED coefficient. This is a good that people will buy more of as income falls, and less of as income rises.

- Example: Bicycle transportation is an inferior good, because Americans demanded MORE bicycles as their incomes fell. If income were to rise, bicycle sales would begin to fall. Since income and quantity move in opposite directions, the YED coefficient for an inferior good is always negative.

Price elasticity of supply and its determinants

- Explain the concept of price elasticity of supply, understanding that it involves responsiveness of quantity supplied to a change in price along a given supply curve.
- Calculate PES using the PES equation.
- Explain, using diagrams and PES values, the concepts of elastic supply, inelastic supply, unit elastic supply, perfectly elastic supply and perfectly inelastic supply.
- Explain the determinants of PES, including time, mobility of factors of production, unused capacity and ability to store stocks.
- Explain why the PES for primary commodities is relatively low and the PES for manufactured products is relatively high.

Price elasticity of supply measures the responsiveness of producers to price changes. Since there is always a direct relationship between price and quantity supplied, the PES coefficient will always be positive. PES can be calculated using the formula:

PES = Percentage change in the quantity supplied ÷ Percentage change in the price

or...

$$\text{PES} = \frac{\% \Delta Q_s}{\% \Delta P}$$

PES will always be positive, since there is a direct relationship between the price of a good and the quantity firms wish to supply.

Consider the following:

- The price of tablet computers rises from \$400 to \$500
- In the week that follows, the quantity rises from 1 million to 1.1 million
- In the three months that follow, the quantity rise from 1 million to 2 million

PES in the short-run (1 week after price change)

$$\begin{aligned} &= ((1.1-1) \div 1) \div ((500-400) \div 400) \\ &= 0.1 \div 0.25 = 0.4 \end{aligned}$$

PES in the long run (3 months after price change)

$$\begin{aligned} &= ((2-1) \div 1) \div ((500-400) \div 400) \\ &= 1 \div 0.25 = 4 \end{aligned}$$

Notice that supply for the same good (tablet computers) is more elastic in the long run than in the short run. This is explained below.

The determinants of PES

The primary **determinant of PES** is the amount of time producers have to respond to a price change.

- In the tablet computer market producers were relatively unresponsive to the rise in price in the one week following the price increase (PES equaled only 0.4)
- After three months, producers had the time to increase their production to meet the higher demand, thus they were much more responsive (PES equaled 4)

Three time periods in determining PES

- The Market Period: Immediately after a change in price. Supply is highly inelastic, because firms cannot immediately produce more of a good.
- The short-run: Firms can use their fixed capital more or less intensively, so supply is more slightly more elastic.
- The long run: Firms have time to vary the amount of capital they use, so supply is highly elastic. In the long run an increase in price will result in a much greater increase in Q_s than in the market period or the short-run.

Other determinants of PES

In addition to the amount of time following a price change, the following help determine PES:

- The mobility of resources: If resources (labor and capital) can be quickly put into or

taken out of the production, supply tends to be more elastic. Generally, this applies to low-skilled manufactured goods, the supply of which is more elastic than high-tech, capital-intensive manufactured goods.

- The ability to store stocks: If large inventories can be kept, producers can respond to price rises by drawing on those inventories to meet rising demand and to price declines by adding to inventories in response to falling demand. Goods that can be stored tend to have more elastic supply than perishable, non-storable goods.

Applications of PES

Similar to PED, knowledge of PES can help businesses and the government better plan for the anticipated price changes to particular goods.

- Business firms: If a producer expects the price of his product to change in the future, he will want to adjust his output accordingly. Being able to adjust output in a timely manner to price changes is key to maximizing a firm's profits.
- Government: A government must consider the PES for a good if it is considering intervening in the market for that good in any way. For example, if a government is considering imposing price controls (maximum or minimum prices) on an agricultural commodity, the PES should be considered so any changes in output resulting from the government controlled price could be anticipated.

Just like PED, the value of PES can be:

- 0-1: Inelastic – Producers are relatively unresponsive to a price change
- 1: Unit Elastic – Producers will respond proportionally to a change in the price
- >1: Elastic – Producers of Good A will be relatively responsive to a price change

1.3 Government Intervention in Markets

Specific and ad valorem taxes

- Explain why governments impose indirect (excise) taxes.
- Distinguish between specific and ad valorem taxes.
- Draw diagrams to show specific and ad valorem taxes, and analyse their impacts on market outcomes.
- Discuss the consequences of imposing an indirect tax on the stakeholders in a market, including consumers, producers and the government.

Up to this point, we have examined how free markets work. A free market is one without any government control or intervention. The price and output is determined by the interactions of buyers and sellers

However, not all markets are completely free. Governments tend to intervene often to influence several variables in markets for particular goods, such as:

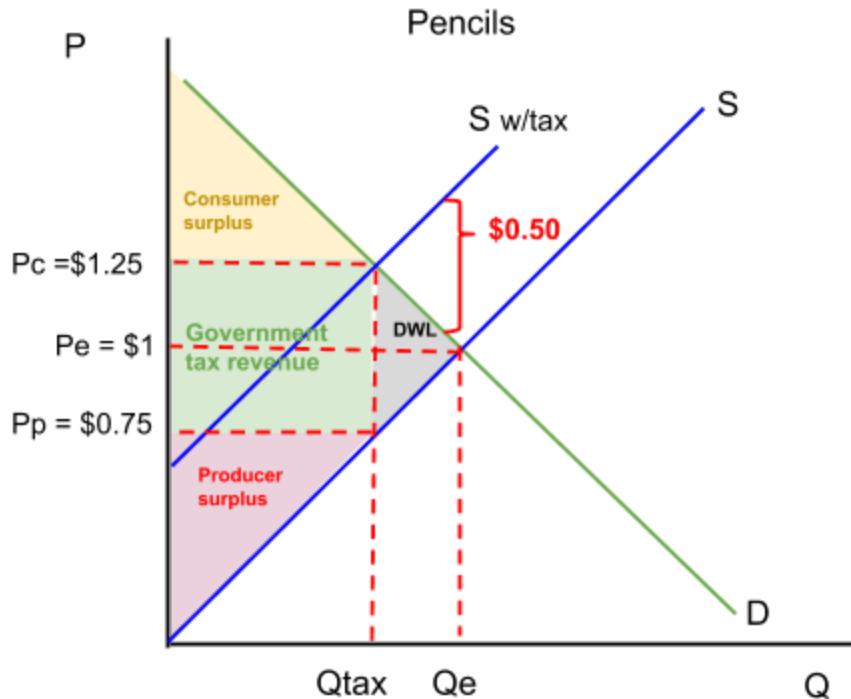
- Taxing the good to discourage consumption or to raise revenues
- Paying producers of the good to reduce costs or to encourage the good's production
- Placing a price ceiling below its free market equilibrium to benefit consumer
- Placing a price floor above its free market equilibrium to benefit producers

When governments intervene in the free market, the level of output and price that results may NOT be the allocatively efficient level. In other words, government intervention may lead to a misallocation of society's resources.

An **indirect tax** is one placed by the government on the producers of a particular good.

- Consumers will pay the tax indirectly through producers
- An indirect tax will be shared by both consumers and producers

For example, assume the government imposes a \$0.50 tax on the production of pencils. The graph below illustrates the effect this tax will have on the market for pencils.



Effect of a \$0.50 per unit tax on pencils:

- The tax is an additional cost for pencil producers, so the supply of pencils decreases.
- Supply will shift UP by \$0.50 (remember, supply also represents marginal cost. The cost of each additional pencil is now \$0.50 higher, so MC shifts up by \$0.50)
- The price of pencils increases from \$1.00 to \$1.25.
- Once the tax is paid, pencil producers get to keep just \$0.75 per pencil

A per unit tax reduces the supply of a good and increases its price. The following points should also be observed.

- Price does not increase by the full amount of the tax
- Producers do not keep the full price paid by consumers, as they must pay the tax
- The tax generates revenue for the government
- There is a loss of total welfare in the market resulting from the tax, and both consumer and producer surplus decrease

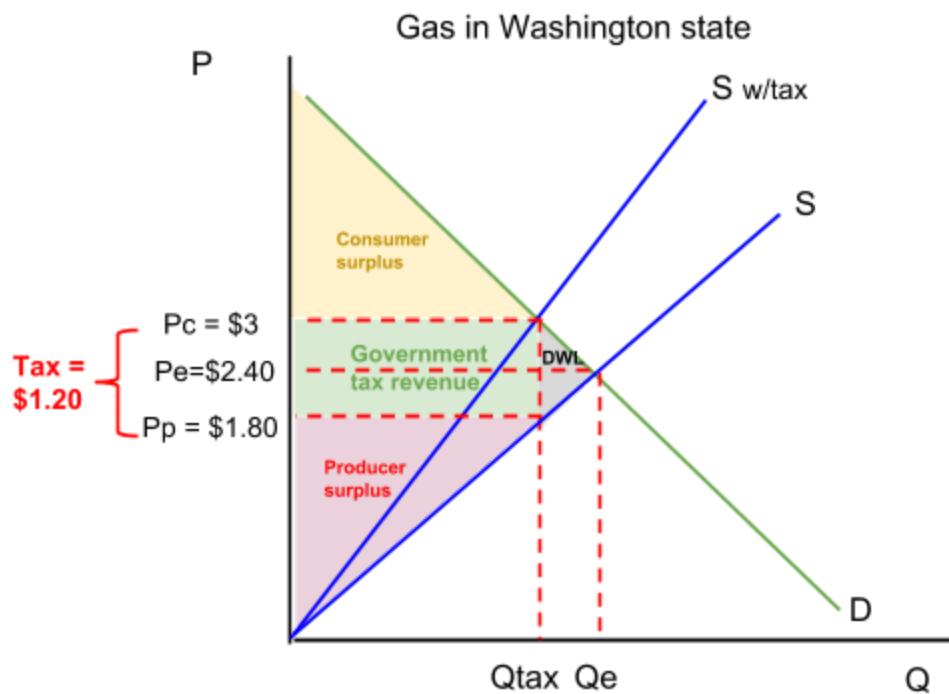
Specific vs. ad valorem taxes

The pencil tax shown above is a **specific tax**, which is a per unit tax of a specific amount, such as \$0.50 per pencil. Another type of per unit tax is an **ad valorem** tax, which is a percentage of the price of a good.

For example, assume the state of Washington charges a 40% tax on each gallon of gasoline sold. Depending on the pre-tax price, the actual dollar amount of the tax will vary. The table below shows the amount of tax charged at a range of pre-tax gas prices.

Pre-tax price per gallon	Amount of tax charged per gallon
\$1	\$0.40
\$2	\$0.80
\$3	\$1.20
\$4	\$1.60

As you can see, an ad valorem tax increases as the good's price increases. Graphically, such a tax will shift the supply curve up vertically by the amount of the tax, just like a specific tax. However, unlike a specific tax, the new supply curve will have a steeper slope than the original curve.



Note from the graph:

- Without any tax, gas in Washington States would cost \$2.40 per gallon.
- The tax shifts the supply curve up by 40% at every price.
- At the \$3 price consumers pay, the government tax a 40% cut (\$1.20), leaving sellers with just \$1.80 per gallon.
- Just like a specific tax, the burden of the tax is shared between consumers and producers.

Tax incidence and price elasticity of demand and supply (HL only)

- Explain, using diagrams, how the incidence of indirect taxes on consumers and firms differs, depending on the price elasticity of demand and on the price elasticity of supply.
- Plot demand and supply curves for a product from linear functions and then illustrate and/or calculate the effects of the imposition of a specific tax on the market (on price, quantity, consumer expenditure, producer revenue, government revenue, consumer surplus and producer surplus).

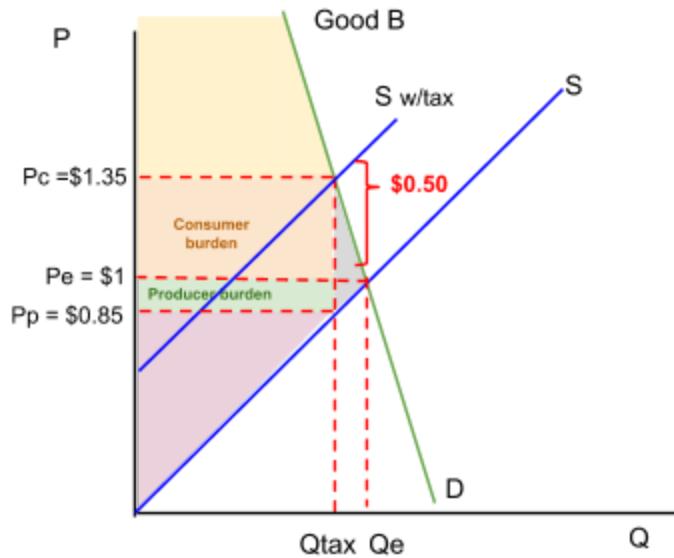
The **incidence** of the pencil tax shown above was equally shared by producers and consumers. Both paid \$0.25 of the \$0.50 tax on each pencil sold. How the burden (incidence) of a tax is distributed between sellers and buyers depends on the price elasticity of demand.

- If demand is relatively elastic, producers will bear the larger burden of the tax. Firms will not be able to raise the price by much out of fear of losing all their customers, therefore price will not increase by much, but producers will get to keep less of what consumers pay.
- If demand is relatively inelastic, consumers will bear the larger burden of the tax. Firms will be able to pass most of the tax onto consumers, who are not very responsive to the higher price, thus will continue to consume close to what they were before the tax.

The implication for government of the above analysis is that if a tax is meant to raise revenue, it is better placed on an inelastic good rather than an elastic good. Taxing elastic goods will reduce the quantity sold and thus not raise much revenue.

Examine the effects of the same \$1 tax on the two goods below, one a highly elastic good, the other a highly inelastic good.





The tax on Good A (highly elastic) is borne mostly by producers, who pass only \$0.15 of the \$0.50 tax onto consumers due to their highly elastic demand.

The tax on Good B (highly inelastic) is borne mostly by consumers. Producers pass \$0.35 of the \$0.50 tax onto consumers due to their highly inelastic demand.

Taxing goods with relatively inelastic demand will raise more revenue and lead to a smaller loss of total surplus, while taxing goods with elastic demand will lead to a larger decrease in quantity and a greater loss of total welfare.

The effects of an indirect tax in linear supply equations (HL only)

A tax is an additional cost placed on producers in a market. Therefore, a tax will affect the supply curve AND the supply equation. One way to think about a tax is that it is a payment made by the producers to the government of a particular amount AFTER consumers have bought the good. Therefore, to show the effect of a tax on a supply equation, we must subtract the amount of the tax from the price consumers paid.

Consider the supply of bread in a small town: $Q_s = -200 + 150P$

Assume a \$1 tax is imposed on bread producers. This means that whatever consumers pay (P), producers will keep \$1 less. The new supply equation is therefore:

$$Q_s = -200 + 150(P - 1)$$

This can be simplified:

$$Q_s = -200 + 150P - 150$$

The new supply of bread is:

$$Q_s = -350 + 150P$$

A \$1 tax on the production of bread cause the supply to decrease. The new supply of bread is:

$$Q_s = -350 + 150P$$

Notice:

- The 'c' variable in the equation decreased. This is the Q-intercept of supply, which is now lower on the Q axis, meaning supply has shifted to the left by 150 units, or up by \$1.
- The 'd' variable has not changed. The tax does not change the responsiveness of producers to price changes. They will still supply 150 more loaves for every \$1 increase in price.

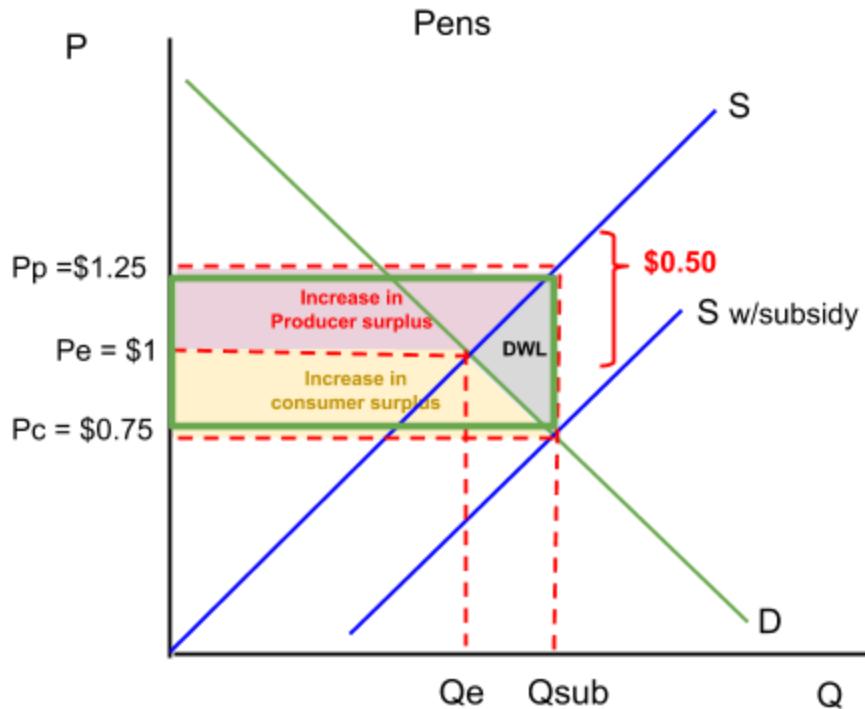
Calculating the effect of a tax on equilibrium: Assume demand for bread is: $Q_d = 600 - 50P$ The tax has no effect on demand, only supply.			
Before the tax: $600 - 50P = -200 + 150P$ $800 = 200P$ $P = \$4$ $Q_d = 600 - 50(4)$ $Q_d = 400$	$600 - 50P = -350 + 150P$ $950 = 200P$ $P = \$4.75$ $Q_d = 600 - 50(4.75)$ $Q_d = 237.5$	After the tax:	

Subsidies

- Explain why governments provide subsidies, and describe examples of subsidies.
- Draw a diagram to show a subsidy, and analyse the impacts of a subsidy on market outcomes.
- Discuss the consequences of providing a subsidy on the stakeholders in a market, including consumers, producers and the government.
- HL only: Plot demand and supply curves for a product from linear functions and then illustrate and/or calculate the effects of the provision of a subsidy on the market (on price, quantity, consumer expenditure, producer revenue, government expenditure, consumer surplus and producer surplus).

A **per unit subsidy** is a payment from the government to producers for each unit produced. Subsidies are sometimes used to encourage a good's consumption by making a good cheaper for consumers.

The effect of a per unit subsidy of \$0.50 on pens is shown below.



Assume the government places a \$0.50 subsidy on the production of pens.

- Supply shifts ‘down’ by \$0.50, since the subsidy reduces the marginal costs of pen producers.
- The price consumers pay falls from \$1 to \$0.75
- The price producers receive for each pen is the \$0.75 consumers pay plus the \$0.50 subsidy, or \$1.25
- There is a greater quantity of pens produced

The effect of the subsidy on consumer and producer surplus is a bit complicated:

- Consumer surplus increases by the yellow area below the original price of \$1 and above the new price of \$0.75
- Producer surplus increase by the red area above \$1 and below \$1.25.
- Consumer surplus and producer surplus now overlap.

The subsidy imposes a cost on the government

- The green rectangle represents the subsidy’s total cost to taxpayers
- The cost is found by multiplying the per unit subsidy by the number of units produced

The subsidy creates a deadweight loss

- The total cost of the subsidy (the green rectangle) is greater than the total benefit of the subsidy (the increases in consumer and producer surplus).
- The burden the subsidy places on taxpayers exceeds the increase in welfare among the buyers and sellers of pens

The effects of a per-unit subsidy in linear supply equations (HL only)

A subsidy is a payment to producers for each unit produced, therefore it reduces the costs of producing each unit of goods. Lower costs increase supply and affect the supply equation

- One way to think about a subsidy is that it is a payment to producers above and beyond the price consumers pay.
- Therefore, to show the effect of a subsidy on a supply equation, we must ADD the amount of the subsidy to the price consumers paid.

Consider the supply of bread in a small town: $Q_s = -200 + 150P$

Assume a \$1 subsidy is provided to bread producers. This means that whatever consumers pay (P), producers will receive \$1 more. The new supply equation is therefore:

$$Q_s = -200 + 150(P+1)$$

This can be simplified:

$$Q_s = -200 + 150P + 150$$

The new supply of bread is:

$$Q_s = -50 + 150P$$

A \$1 subsidy to the producers of bread causes the supply to increase. The new supply of bread is:

$$Q_s = -50 + 150P$$

Notice:

- The 'c' variable in the equation increased. This is the Q-intercept of supply, which is now closer to the origin on the Q axis, meaning supply has shifted to the right by 150 units
- The 'd' variable has not changed. The subsidy does not change the responsiveness of producers to price changes. They will still supply 150 more loaves for every \$1 price increase

Calculating the effect of a tax on equilibrium: Assume demand for bread is:

$$Q_d = 600 - 50P$$

The tax has no effect on demand, only supply.

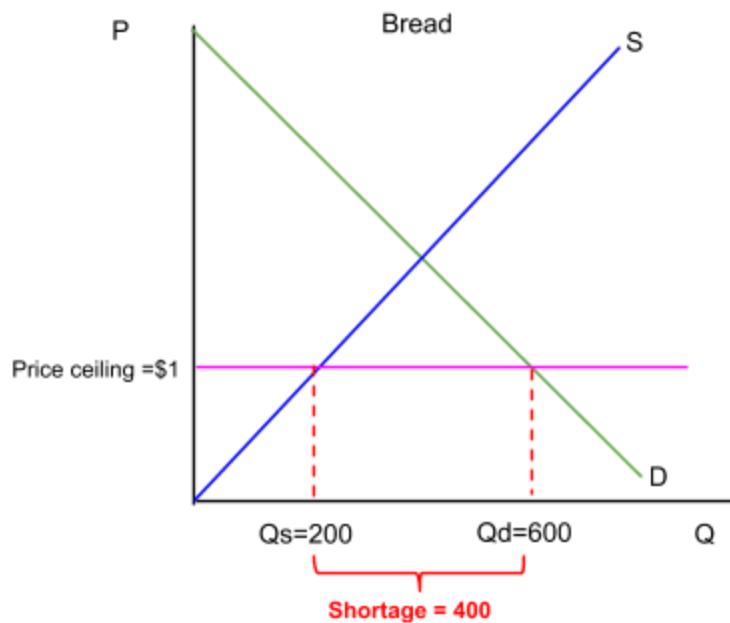
Before the subsidy:	$600 - 50P = -200 + 150P$ $800 = 200P$ $P = \$4$ $Q_d = 600 - 50(4)$ $Q_d = 400$	After the subsidy:	$600 - 50P = -50 + 150P$ $650 = 200P$ $P = \$3.25$ $Q_d = 600 - 50(3.25)$ $Q_d = 437.5$
---------------------	--	--------------------	---

Price ceilings

- Explain why governments impose price ceilings, and describe examples of price ceilings, including food price controls and rent controls.
- Draw a diagram to show a price ceiling, and analyse the impacts of a price ceiling on market outcomes.
- Examine the possible consequences of a price ceiling, including shortages, inefficient resource allocation, welfare impacts, underground parallel markets and non-price rationing mechanisms.
- Discuss the consequences of imposing a price ceiling on the stakeholders in a market, including consumers, producers and the government.
- HL only: Calculate possible effects from the price ceiling diagram, including the resulting shortage and the change in consumer expenditure (which is equal to the change in firm revenue).

A **price ceiling** is a maximum price set below the equilibrium price meant to help consumers of a product by keeping the price low.

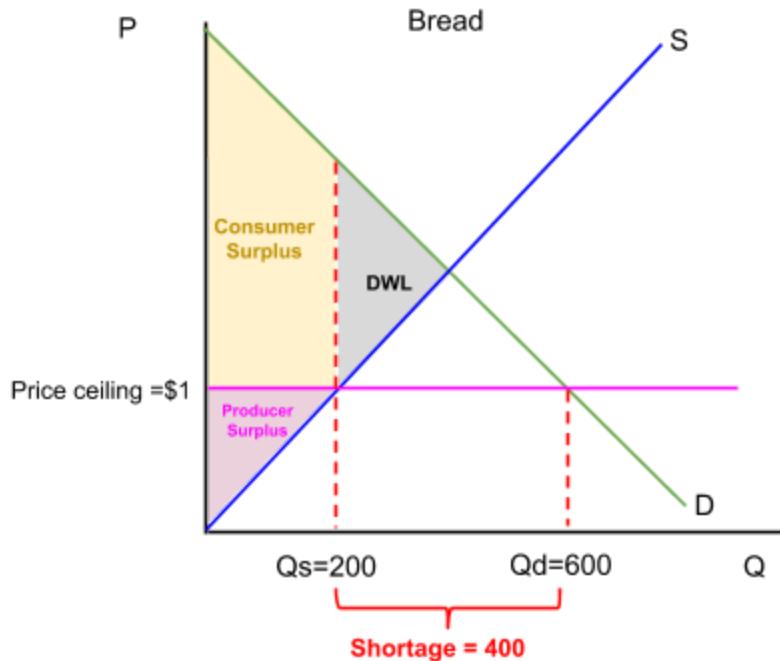
Assume the government wishes to keep bread, a staple food, cheap for consumers. A price ceiling set below the equilibrium price will result in cheaper bread, but also cause a disequilibrium in the bread market.



The price ceiling has led bakeries to reduce their quantity supplied and consumers to increase their quantity demanded, leading to a shortage of bread.

When a government places a **binding price ceiling** on a good to help consumers, there are several effects that we can observe in the market. Assume the government has intervened in

the market for gasoline to make transportation more affordable for the nation's households



Effect on consumers:

- Quantity demanded increases (Q_d)
- The lower price leads to an increase in consumer surplus
- The lower quantity means some consumers who want to will not be able to buy the good

Effect on producers:

- Quantity supply decreases (Q_s)
- The lower price means less producer surplus
- The lower quantity means some producers will have to leave the market and output will decline

Effect on the market:

- Overall, not enough bread is produced, and the market is allocatively inefficient. The gray triangle represents the loss of total welfare resulting from the price ceiling (Deadweight Loss)

Price floors

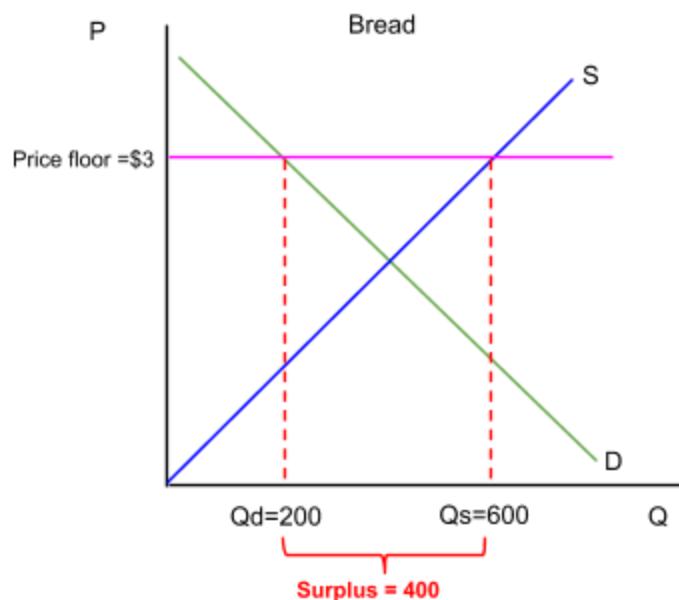
- Explain why governments impose price floors, and describe examples of price floors, including price support for agricultural products and minimum wages.
- Draw a diagram of a price floor, and analyse the impacts of a price floor on market outcomes.
- Examine the possible consequences of a price floor, including surpluses and government measures to dispose of the surpluses, inefficient resource allocation and welfare impacts.
- Discuss the consequences of imposing a price floor on the stakeholders in a market,

- including consumers, producers and the government.
- HL only: Calculate possible effects from the price floor diagram, including the resulting surplus, the change in consumer expenditure, the change in producer revenue, and government expenditure to purchase the surplus.

What if instead of helping buyers, the government decides it needs to help bakery owners. Instead of a price ceiling, the government can set a **price floor**.

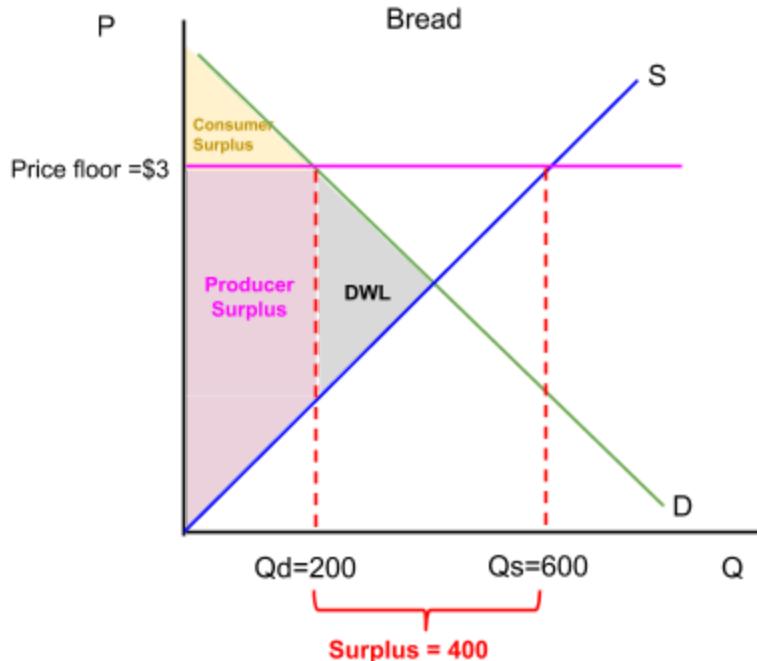
A **price floor** is a minimum price set above equilibrium meant to help producers in a market.

A price floor set above the equilibrium price of \$2 in the bread market will cause a disequilibrium in the market.



The higher price has led bakeries to produce more bread, but it has caused the quantity demanded to decrease. There is now a surplus of bread on the market. Without the price floor, the excess supply would be eliminated as the price returned to equilibrium. However, there is no legal way for bakers to sell their bread in the presence of the minimum price.

When a government places a **binding price floor** on a good to help producers, there are several effects that we can observe in the market. Assume the government has intervened in the market for corn to help farmers sell their crop at a price that allows them to earn a small profit.



Effect on consumers:

- Quantity demanded decreases (Q_d)
- The higher price means there is less consumer surplus

Effect on Producers:

- Quantity supplied increases (Q_s)
- The higher price means there is more producer surplus
- Since consumers only demand Q_d , there is an excess supply of unsold bread ($Q_s - Q_d$)

Effect on the market:

- Overall, the market produces too much bread and is thus allocatively inefficient. The increase in producer surplus is smaller than the decrease in consumer surplus. The total loss of welfare is represented by the gray triangle.

Price controls and black markets

Markets in which binding price controls are creating shortages or surpluses tend to result in **black markets** in which the shortage or surplus is resolved.

- In the case of a price ceiling, the black market price will be higher as those who were unable to find the good in the formal market will turn to the informal, black market to acquire the needed goods at a higher than legal price.
- In the case of a price floor, the black market price will be lower as sellers who were unable to sell their surplus output will turn to the informal, black market to sell their excess supply at a lower than legal price.

Calculating the effects of price controls using linear equations (HL only)

As with taxes and subsidies, we can use linear supply and demand equations to calculate the

effects of price ceilings and price floors. Once again, assume demand and supply for bread is:

$$Q_d = 600 - 50P \text{ and } Q_s = -200 + 150P$$

As we have already shown, the current equilibrium price is \$4 and the quantity is 400 loaves. Assume the government wishes to help households afford bread, so imposes a price ceiling of \$3 on bread. To determine the impact on the market, we must simply put \$3 into both equations.

$$\begin{aligned} Q_d &= 600 - 50(3) = 450 \text{ loaves} \\ Q_s &= -200 + 150(3) = 250 \text{ loaves} \end{aligned}$$

The \$3.00 price ceiling will create a shortage of 200 loaves of bread.

- Producers will reduce their output of bread and more consumers will wish to buy bread.
- The price ceiling took a market that was efficient and made it inefficient. Not enough resources are allocated towards bread production as a result of the price ceiling

Next, assume that the government determines that \$4 is too cheap for bread, and producers need the price to be higher. The government imposes a price floor of \$5 in the market. To determine the impact on the market, we must simply put \$5 into both equations.

$$\begin{aligned} Q_d &= 600 - 50(5) = 350 \text{ loaves} \\ Q_s &= -200 + 150(5) = 550 \text{ loaves} \end{aligned}$$

The \$5.00 price floor will create a surplus of 200 loaves of bread.

- Producers will increase their production of bread to take advantage of the now higher prices it is commanding in the market.
- Consumers will reduce the quantity of bread they demand due to the now higher price.
- The price floor took an efficient market and made it allocatively inefficient. Too many resources are now being allocated towards bread production!

Conclusion: Price controls rarely increase efficiency or total welfare in a market. They result in either shortages (price ceilings) or surpluses (price floors), and therefore lead to a net loss in total welfare for society. Some benefit, but many suffer.

1.4 Market Failure

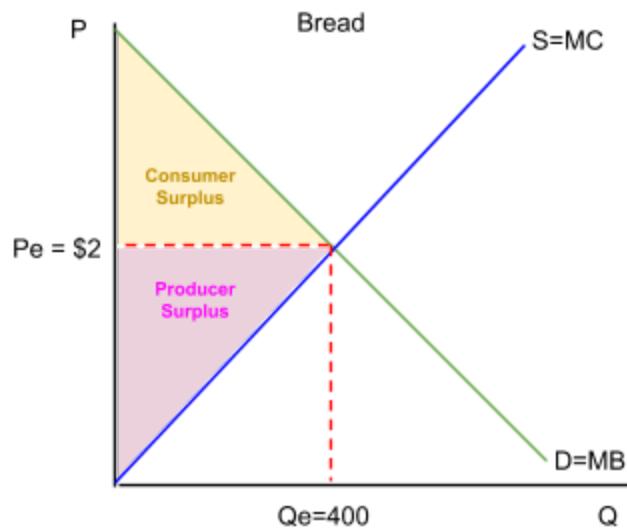
The meaning of market failure

- Examine the concept of market failure as a failure of the market to achieve allocative efficiency, resulting in an over-allocation of resources (over-provision of a good) or an under-allocation of resources (under-provision of a good)

Revisiting allocative efficiency

The optimal quantity of a good occurs when the marginal benefit of consuming the last unit equals the marginal cost of producing that last unit. At the $MB=MC$ level of output, total economic surplus is maximized.

We have learned that in competitive markets, in which there are no barriers to entry and exit, in which there are a large number of firms producing a product, and in which there are no negative effects from the good's production or consumption, a socially optimal level of output will be produced. The graph below shows just such a market.



In the competitive bread market above, the price of \$2 results in an efficient market outcome:

- The quantity demanded equals the quantity supplied (no shortages or surpluses exist)
- The marginal benefit of the last loaf of bread is equal to the marginal cost of producing it
- Total economic surplus is maximized. There is no other price/quantity combination that would result in a greater level of consumer and producer surplus than \$4 and 400 loaves of bread.

Inefficient market outcomes

While competitive markets in which no harmful effects arise from the good's production or consumption will be efficient, that leads us to ask the important question: When is a market

inefficient? Inefficiency results from several possible situations, including:

- The existence of **externalities**: Any time there are spillover costs or benefits passed onto a third party not involved in a market, which result from the good's production or consumption, the equilibrium quantity will be greater than or less than the socially optimal quantity.
- The existence of **public goods**: Goods that are non-rivalrous in their consumption and non-excludable by producers will be under-provided by the free market. In extreme cases, some goods that society needs will simply not be provided at all by private producers, resulting in a large inefficiency.
- The existence of **asymmetric information**: If the sellers of a good withhold certain information about the good from the buyers, or vice versa, then the market will produce the good at an inefficient level.
- The existence of **imperfect competition**: Any time a market is imperfectly competitive the equilibrium quantity will be lower and the price will be higher than the socially optimal quantity and price.

The meaning of externalities

- Explain the concepts of marginal private benefits (MPB), marginal social benefits (MSB), marginal private costs (MPC) and marginal social costs (MSC).
- Describe the meaning of externalities as the failure of the market to achieve a social optimum where $MSB = MSC$.

An **externality** exists anytime the production or consumption of a good creates spillover benefits or costs on a third party not involved in the market. In such cases, resources will either be under-allocated (positive externalities) or over-allocated (negative externalities) towards the production of certain goods.

Examples of Positive Externalities (known as merit goods)	Examples of Negative Externalities (known as demerit goods)
Receiving a college education makes the consumer more likely to contribute to the well-being of society as a whole	Driving sports-utility vehicles contributes to traffic and contributes more to global warming
Riding bicycles to work reduces congestion on the roads and makes for less traffic for everyone else	Producing electricity using coal creates greenhouse gas emissions and air pollution
Getting vaccines against communicable diseases reduces the chance you will get others sick	Smoking cigarettes contributes to lung disease among not just the smokers, but those who suffer from second-hand smoke

Consumers and producers make decisions based on their private benefits and costs. If the equilibrium achieved at the intersection of supply (marginal private cost, or MPC) and demand (marginal private benefit, or MPB) is either greater than or less than the socially optimal equilibrium (where marginal social benefit, or MSB, equals marginal social cost, or

MSC), then an externality exists.

Socially optimal equilibrium is where $MSC=MSB$

Free market equilibrium is where $MPC=MPB$

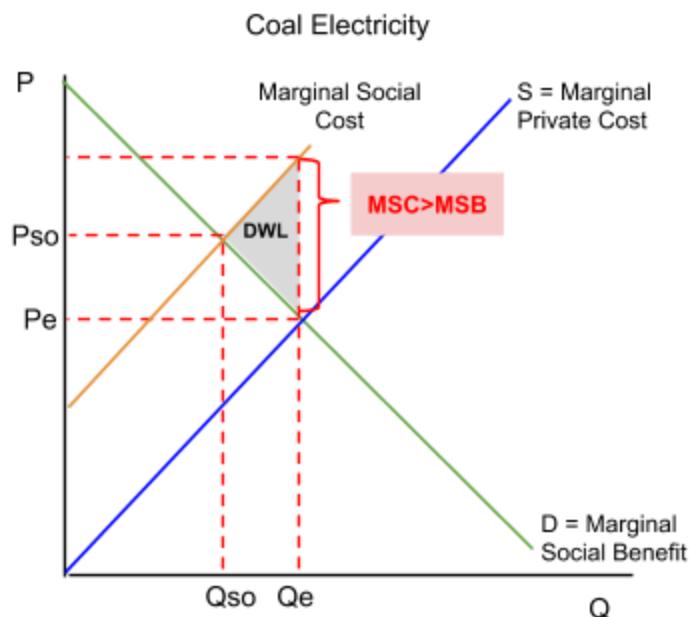
If $MPC \neq MSC$ or if $MPB \neq MSB$
then an externality exists and a market failure will result.

Negative externalities of production and consumption

- Explain, using diagrams and examples, the concepts of negative externalities of production and consumption, and the welfare loss associated with the production or consumption of a good or service.
- Explain that demerit goods are goods whose consumption creates external costs.
- Evaluate, using diagrams, the use of policy responses, including market-based policies (taxation and tradable permits), and government regulations, to the problem of negative externalities of production and consumption

Negative externalities of production arise when the production of a good creates spillover costs on a third party, which is often times the environment as a whole. The Marginal Social Cost of producing a good is greater than the Marginal Private Cost of producing it.

For example, in the market for coal electricity, the production of electricity creates pollution that can harm human health and the environment. The graph below shows the effect of these negative environmental production externalities in the market for coal electricity.



The labels in the graph above are described below:

- MPC = The marginal private costs of producing electricity by burning coal

- MSC = the cost to society of producing electricity by burning coal. MSC = the MPC + external costs
- MSB = the marginal benefit to society of coal electricity
- Q_e = the actual output in the market
- Q_{so} = the socially optimal output in the market, where $MSC=MSB$
- P_e = the equilibrium price in the market
- P_{so} = the socially optimal price if all social costs were considered in the good's production. P_{so} is where $MSB=MSC$

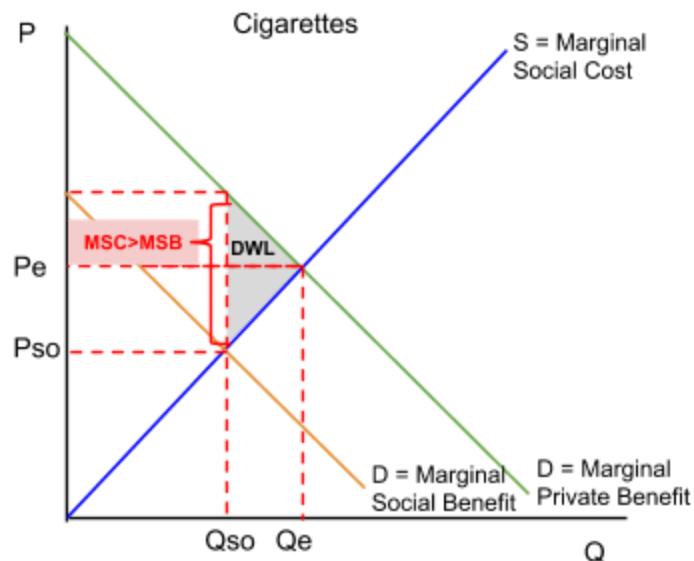
A polluting industry creates costs for society that are not paid by the polluting firm. These external costs of production may include:

- Greenhouse gas emissions which contribute to global warming
- Air pollution
- Contributions to lung disease and cancer rates among the population
- Water pollution which destroys fish stocks
- Soil contamination that harms agricultural productivity.

The existence of all these externalities creates a social cost that exceeds the private cost!

- As a result, there is a deadweight loss created by the unregulated industry
- At the equilibrium output of Q_e , the marginal social cost exceeds the marginal social benefit, meaning too much of the good is being produced by the free market! This is a market failure!

Some goods are over-consumed by the free market. This would be the case if the process of consuming a good creates spillover costs on a third party. The classic example of a **negative externality of consumption** is the second-hand smoke effects of smoking cigarettes.



In the market for cigarettes we can observe the following:

- The Marginal Private Benefit (MPB) of smoking cigarettes is greater than the

Marginal Social Benefit (MSB)

- Smoking creates costs (negative benefits) on non-smokers, so society benefits as a whole less than the smokers themselves
- There are no externalities in the production of cigarettes, so the supply curve represents the private costs and the social costs.
- The equilibrium price (P_e) is greater than the price would if demand represented the social benefits of smoking (P_{so})
- The equilibrium quantity (Q_e) is greater than the socially optimal quantity (Q_{so})

Smoking harms third parties who do not buy or sell cigarettes, therefore the cigarette market results in a negative consumption externality.

Notice on the graph:

- At Q_e (the actual quantity of cigarettes consumed in a free market), the MSC of smoking exceeds the MSB.
- Too many cigarettes are being produced and consumed at Q_e , resulting in a loss of total welfare equal to the gray triangle.

Resources will be over-allocated towards the production and consumption of cigarettes by the free market. This is a market failure!

Government responses to negative externalities

Whenever a market fails by allocating too many resources towards the production of the good, the government can potentially improve market efficiency by intervening to reduce the quantity produced and consumed to a more socially optimal level (where $MSB=MSC$).

Methods a government might take to correct a negative externality include:

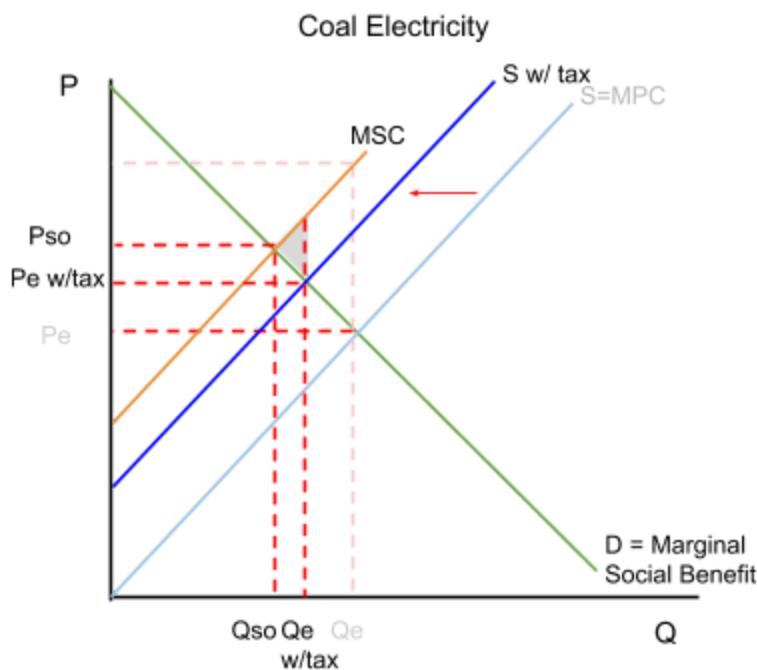
- **Corrective taxes:** A per-unit tax on a good meant to reduce the supply, increase the price, and reduce the quantity demanded to a more socially optimal level. Unlike a tax on a good that is produced efficiently by the free market, a corrective tax is meant to correct a failing market and help the market achieve a higher level of efficiency.
- **Regulation/legislation:** Laws that limit the quantity of a good produced or require it to be produced in an environmentally friendly way may increase the costs of production to firms and reduce the quantity to a more socially optimal level.
- **Banning:** Many goods that create negative externalities are simply banned. Examples include: Drinking among minors, narcotics, prostitution, fully automatic weapons, etc.
- **Tradable Permits:** Issuing permits to producers of goods that create negative environmental externalities will create a physical limit on the amount of pollution or of the harmful activity, reducing the overall cost to society of the activity.

Government responses to negative externalities – corrective taxes

A tax meant to correct a market failure is sometimes referred to as a Pigouvian Tax, after the economist Arthur Pigou, who first proposed using taxes to reduce the output of harmful goods.

- Recall that a tax is a determinant of supply
- A tax on a good that created externalities of production or consumption will increase the marginal private costs of production and reduce the supply to a level closer to the marginal social costs (which include all external costs).

The tax reduces the supply, causing the equilibrium price to rise and the quantity demanded to fall. If the size of the tax reflects the size of the external costs, then the new equilibrium output (Q_e) will equal the socially optimal level of output (Q_{so}). Observe the effect of a tax on coal electricity producers below.



The effects of a tax on a good with a negative externality:

- The tax has reduced supply because it increases the marginal private cost of producing coal electricity
- Equilibrium price has increased from P_e to $P_{e \text{ w/tax}}$
- Equilibrium quantity has decreased from Q_e to $Q_{e \text{ w/tax}}$; less coal electricity is demanded
- The amount of deadweight loss has decreased because less coal is being used to make electricity
- The Pigouvian tax has reduces the loss of total surplus created by the negative externalities of production in the market

Corrective taxes are a popular response to negative externalities among economist, but among policy-makers, they are rarely popular. Some arguments against corrective taxes include:

- Higher costs for producers: Producers face higher costs, and therefore will reduce their output of the goods being taxed. This is bad for business.

- Higher prices for consumers: Consumers of the goods being taxed face higher prices, reducing consumer surplus and the real incomes of households. For some goods (such as electricity) this could place a major financial burden on households.
- Less employment: As the taxed industries reduce their output, they may be forced to lay off workers, increasing unemployment in the economy.
- Loss of competitiveness in global market: This is a major one. Policy-makers fear that if they impose taxes on their nation's producers, but other nations' governments do not impose taxes on their producers, then the domestic industries will suffer while foreign producers thrive. International cooperation on the implementation of corrective taxes could eliminate this problem, but there have been very few examples of such cooperation.

Government responses to negative externalities – tradable permits

A second method for reducing the negative externalities arising from production or consumption of certain goods is the use of **tradable permits**. For example, in Europe there is a market for permits to emit carbon dioxide, a greenhouse gas widely believed to contribute to global warming. Here's how it works:

1. A government or multi-national governing body issues or auctions off permits to polluting industries that allow them to emit a certain amount of carbon.
2. Some firms pollute beyond their permitted amount, so will either have to acquire more permits or reduce their emissions
3. To acquire more permits, they must buy them from in the market from firms that do not need all of their permits
4. The supply of permits is fixed and determined by the government, the demand for permits therefore determines the price of pollution. The more firms want to pollute, the more expensive it becomes to pollute.
5. There is a strong incentive for firms to reduce their emissions, because they can then sell the permits they do not need, adding to firm profit.

A tradable permit scheme has several advantages over corrective taxes, and some disadvantages...

Advantages of tradable permits:

- Creates a strong incentive to reduce pollution, since permits can be sold off for profit
- Creates a clear price for pollution, internalizing the costs that firms would have externalized without the scheme.
- Places a clear limit on the quantity of pollution that will be created each year
- Price of permits can be increased over time by reducing the number of permits available.

Disadvantages of tradable permits:

- The price of permits is determined by the free market, and may be too low to create strong incentives to reduce pollution
- The amount of permits is decided by government, and may be too high if polluting

- industries are allowed to influence policy
- It is costly and difficult to monitor industries to make sure everyone who pollutes has the permits to do so.

Government responses to negative externalities – regulation

Regulation of polluting or harmful industries is another option for governments to attempt and promote a more socially optimal level of output of a demerit good. Some factors to consider when a government regulates an industry include:

- Monitoring: The government must monitor emissions of polluters, which can be costly and difficult.
- Enforcement: The government must have a way to enforce legislation on polluters.
- Penalties: The penalties for violations must be significant enough to dissuade firms from ignoring legislation
- Incentives: If the penalty is not harsh enough, the firm will simply ignore regulations and pollute anyway. The fine must be greater than the cost of pollution abatement, otherwise firms will keep polluting.

The effect of regulation is similar to a tax or the requirement that firms must buy permits for pollution, regulation will add to the cost of producing harmful goods. Firms face higher costs in adhering to regulations, reducing the supply of demerit goods and creating incentives for firms to produce goods in more environmentally and socially responsible ways.

The intended effect of government regulations of externalizing industries is to force the polluters to incur costs associated with pollution control. Firms forced to reduce their pollution will face higher costs, shifting the market supply curve for a polluting product to the left. Equilibrium quantity should fall closer to the socially optimal level.

Positive externalities of production and consumption

- Explain, using diagrams and examples, the concepts of positive externalities of production and consumption, and the welfare loss associated with the production or consumption of a good or service.
- Explain that merit goods are goods whose consumption creates external benefits.
- Evaluate, using diagrams, the use of government responses, including subsidies, legislation, advertising to influence behaviour, and direct provision of goods and services.

A **positive externality of production** exists if the production of a good or service provides spillover benefits to a third party not involved in the market. For example, consider the market for ecotourism .

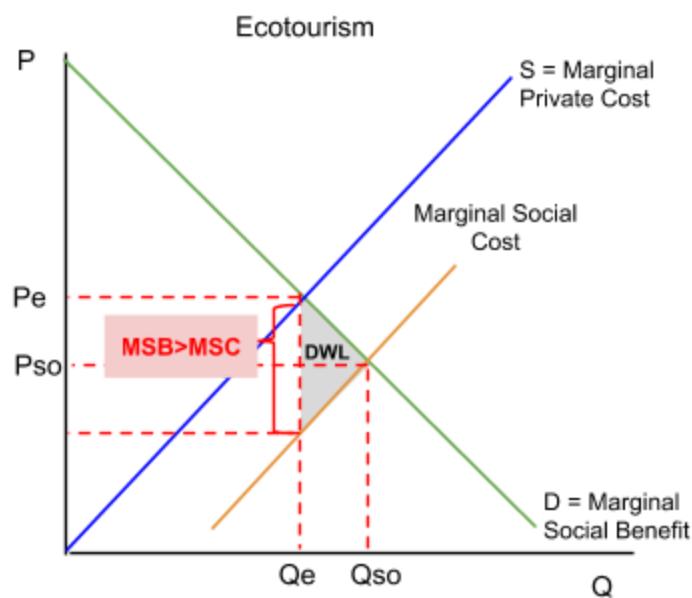
- In many parts of the world, including in the Amazon rainforest, in Costa Rica, in Malaysian Borneo, Western Canada and elsewhere, a large eco-tourism industry has developed.
- This industry provides paying customers with the experience of outdoor adventures in nature

- But the existence of this industry creates positive benefits for society beyond those who pay for the experience.

Positive externalities of ecotourism include:

- Forests left standing act as a “carbon sink”, absorbing CO₂ emitted from the production of other consumer goods.
- Protection of ecosystems that might otherwise be exploited or developed
- Wildlife populations may remain protected and intact whereas they otherwise may dwindle due to habitat destruction and over-hunting
- Water resources (rivers, lakes) are protected, allowing downstream users to benefit from clean water for cooking, cleaning, drinking, etc.

The existence of a positive production externality can be illustrated graphically as a market in which the Marginal Private Cost of production (MPC) is greater than the Marginal Social Cost of production (MSC). As a result, the free market will provide a quantity of the good that is less than the socially optimal quantity.



In the market for ecotourism:

- Operating a business in the industry is expensive, so the MPC is relatively high.
- There are external benefits of operating an ecotourism business, which are reflected in the lower MSC.
- The equilibrium price (P_e) is higher than what is socially optimal (P_{so}). The quantity demanded would be greater for ecotourism if the price were lower.
- The equilibrium quantity (Q_e) is less than what is socially optimal (Q_{so}). Society would be better off with more businesses offering eco-tourism services.

Because the free market will provide not a socially optimal amount of ecotourism services, there is an area of deadweight loss in the market diagram. If a greater quantity of the merit

good were produced and consumed, society as a whole would be better off.

At the equilibrium quantity and price:

- The MSB is greater than the MSC. Society benefits more than it costs to provide Q_e , resources are under-allocated towards eco-tourism
- Society stands to gain an amount of welfare equal to the gray triangle if more eco-tourism can be provided.
- The price (P_e) is too high, and therefore the equilibrium quantity demanded (Q_e) too low.

Increased provision of merit goods like eco-tourism would benefit society as a whole

Positive externalities of consumption

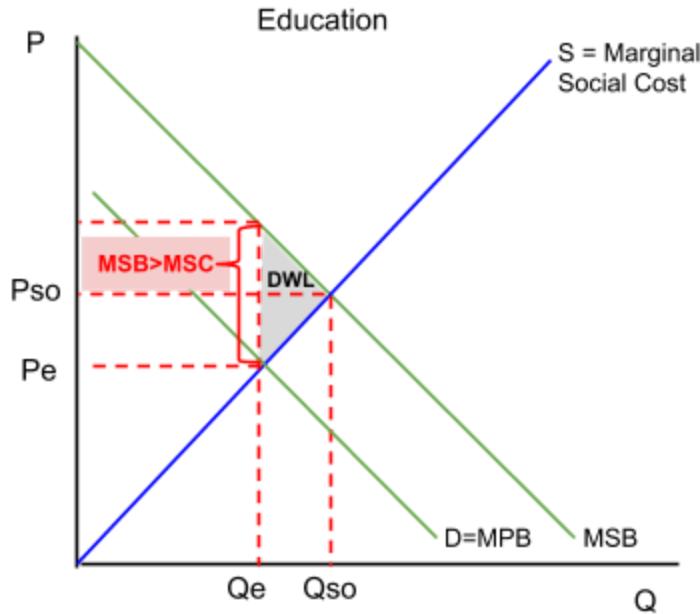
A **positive externality of consumption** exists if the consumption of a good or service provides spillover benefits to a third party not involved in the market. For example, consider the market for education.

- Getting an education provides many benefits for the student, such as better job opportunities, higher pay, an earlier retirement and better travel opportunities
- However, receiving an education also benefits society as a whole.

Positive externalities of education include:

- An educated citizen will be more productive in his or her life, contributing more to national output,
- He or she will pay more in taxes, which go towards providing benefits for everyone in society, even those without an education.
- He or she is more likely to become a business owner, offering employment opportunities to others in society that may not otherwise have been provided.

Education is a merit good that provides spillover benefits to society as a whole. In the market, the social benefits exceed the private benefits of receiving an education, as seen in the graph below.



In the market for education:

- Private demand for education is equal to the marginal private benefit.
- The quantity of education society will consume if it is left entirely to the free market is Q_e , but this is less than what is socially optimal (Q_{so})
- There are external benefits of receiving an education, represented by the vertical distance between MPB and MSB

Because there are external benefits of consuming education:

- The Marginal Social Benefits of receiving an education (MSB) are greater than the Marginal Private Benefits of receiving an education (MPB).
- If left to the free market, too few people will receive an education

Anytime the free market provides too little of a good or service, there is a potential gain in total surplus of the good being produced at a greater quantity.

At the equilibrium price and quantity:

- The marginal social benefit of education is greater than the marginal social cost, indicating that not enough education is being provided.
- At the socially optimal quantity (Q_{so}), the $MSB=MSC$, indicating that this is the allocatively efficient level of education to provide.
- If demand were greater, the price would be higher and more institutions would provide education, increasing the quantity supplied.
- Increased provision of merit goods like education would benefit society as a whole.

Government responses to positive externalities

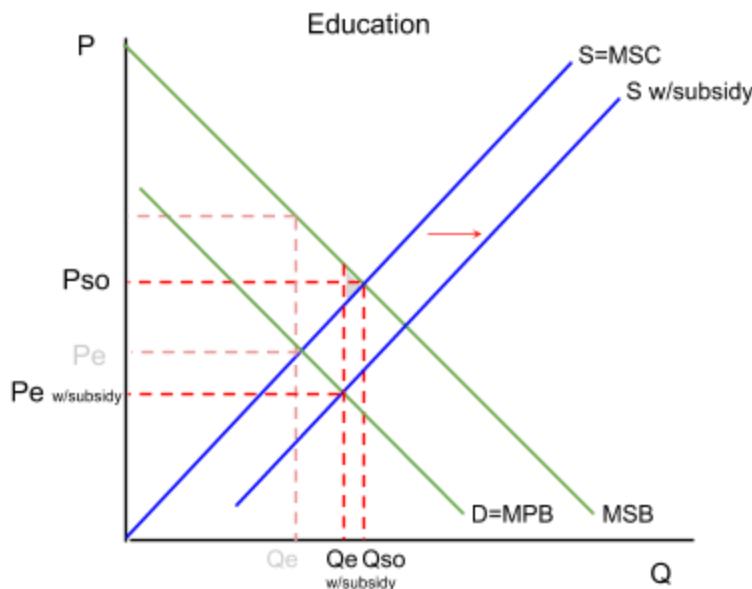
When a market fails by under-allocating resources towards the production of a good, society stands to benefit from increasing the production and consumption of the good in the

market. Therefore, government policies aimed at increasing either the supply or the demand for the good can improve efficiency in the market for merit goods. Such policies include:

- Corrective subsidies (to producers): A subsidy is a payment from the government to producers. Subsidies lower the marginal private costs of production, increasing the supply, reducing the price and increasing the quantity demanded for the good being subsidized.
- Corrective subsidies (to consumers): A subsidy to consumers of a good will increase the marginal private benefit of consumption (since individuals now get paid to buy a good) and increase the demand for the good. The higher price incentivizes firms to provide a greater quantity, resulting in a more efficient allocation of resources towards the good
- Government provision: Many merit goods are provided by the government, such as education, health care, infrastructure like bridges and airports, police security, and so on.
- Positive advertising: Government programs that educate consumers about the positive private and social benefits of a good may increase demand for the good, incentivizing firms to produce more of it. Examples include healthy eating campaigns, safe sex campaigns (to encourage condom use) promoting flu shots (and other vaccines), and so on...

Corrective subsidies

A subsidy to producers reduces the marginal private costs of production and increases the supply of the good being subsidized. In the market below, the government is subsidizing education, reducing the private costs of providing an education and increasing the quantity demanded.



The effects of a subsidy on a good that creates positive externalities:

- The subsidy has lowered the private costs of providing an education

- The supply of education increases, reducing the price and increasing the quantity demanded
- More students can afford an education, increasing the equilibrium quantity closer to the socially optimal level
- The area of deadweight loss is reduced due to government spending on education

Lack of public goods

- Using the concepts of rivalry and excludability, and providing examples, distinguish between public goods (non-rivalrous and non excludable) and private goods (rivalrous and excludable).
- Explain, with reference to the free rider problem, how the lack of public goods indicates market failure.
- Discuss the implications of the direct provision of public goods by government.

So far we have heard about markets failing when they:

- Produce too much of a good (negative externalities)
- Produce too little of a good (positive externalities)

But what if a market fails to produce any amount of a good that creates social benefits? A good that is not provided by the free market at all is known as a public good.

A **public good** is any good that provides benefits to society which are non-rivalrous, and the good is non-excludable by the producer. Because of these characteristics, public goods will not be provided by the free market at all.

To be considered public, a good must be:

- **Non-rivalrous** in consumption: This means that one consumer's enjoyment of the benefits of a good does not diminish any other consumer's enjoyment of its benefits.
- **Non-excludable** by the provider: This means that once a good has been provided, it is not possible to exclude any individuals from enjoying its benefits. In other words, you can't make individuals pay for the good once it is made available. There will be free-riders, or individuals who enjoy the good's benefits without ever paying for it!

Examples of public goods

To find examples of public goods, all we have to do is walk out our front door and look around. Many of the goods and services a government provides its citizens using tax money are examples of public goods. These may include:

- Infrastructure: Roads, sidewalks, street lights, power lines, sewage systems, train tracks... many of these goods are non-excludable and non-rivalrous, therefore are unlikely to be provided by the free market. Government must provide such goods so that society can enjoy their benefits.
- Parks: Think of the last time you walked through a public park. Did you have to pay to get in? (If not, then it was non-excludable). Did your enjoyment of the park prevent others from enjoying it? (If not, then it was non-rivalrous). Public parks are an example of public goods.

- Fire and Police Protection: If your house catches on fire, do you have to call a private fire-fighting firm to come put it out? The reason you don't is because the benefits of having fire protection are non-rivalrous. Putting the fire in your house out will benefit your neighbors, whose houses are less likely to burn down. Police protection is the same way. Without government-provided police force, society as a whole would be unsafe because very few people would choose to hire private security. The benefits of police protection are non-rivalrous and non-excludable.
- National Defense: An army, navy and air force provide citizens with protection, which, once provided, individuals within the nation cannot be excluded from benefiting from. One person's safety does not diminish others', so defense is non-excludable and non-rivalrous: a purely public good.

Common access resources and the threat to sustainability

- Explain, using examples, common access resources.
- Apply the concept of sustainability to the problem of common access resources.
- Examine the consequences of the lack of a pricing mechanism for common access resources in terms of goods being overused/ depleted/degraded as a result of activities of producers and consumers who do not pay for the resources that they use, and that this poses a threat to sustainability.
- Discuss, using negative externalities diagrams, the view that economic activity requiring the use of fossil fuels to satisfy demand poses a threat to sustainability.
- Discuss the view that the existence of poverty in economically less developed countries creates negative externalities through overexploitation of land for agriculture, and that this poses a threat to sustainability.
- Evaluate, using diagrams, possible government responses to threats to sustainability, including legislation, carbon taxes, cap and trade schemes, and funding for clean technologies.
- Explain, using examples, that government responses to threats to sustainability are limited by the global nature of the problems and the lack of ownership of common access resources, and that effective responses require international cooperation

In addition to merit goods, demerit goods and public goods, a third type of market failure arises from the existence of **common access resources**.

Common access resources are those "gifts of nature" over which there is no private ownership, and therefore no effective means of regulating the use of the resource.

Examples of common access resources include:

- Fish in the sea
- Trees in a forest
- Common pastureland
- Fresh water in aquifers or in rivers

In each of these cases, the lack of ownership over the resources creates an incentive for potential users to exploit them to the fullest extent possible, so as to extract as much benefit as possible before other users extract and exploit the resource. This is known as the **tragedy**

of the commons.

Possible solutions to the tragedy of the commons

When there exist a common resource, for which there is no private owner, the incentive among rational users of that resources is to exploit it to the fullest potential in order to maximize their own self gain before the resource is depleted.

- The tragedy of the commons, therefore, is that common resources will inevitably be depleted due to humans' self-interested behavior, leaving us with shortages in key resources essential to human survival.
- This represents a market failure because, without allocation of property rights over or effective management of common access resources, they will be exploited unsustainably

Sustainability is defined as the ability of an activity or a resource to endure for the use and enjoyment of future generations

Possible Solutions to the Tragedy of the Commons:	
Privatization:	Assigning private ownership over a resource creates an incentive among the private owners to protect and manage its use in a sustainable manner, so as to benefit from its existence into the future.
Government management:	Strict government control over the access to and use of common resources may limit access to them to a sustainable level.
Tradable permits:	Issuing permits to private users to allow a certain amount of extraction in a period of time may limit the exploitation of the resource to sustainable level.

Asymmetric information (HL only)

- Explain, using examples, that market failure may occur when one party in an economic transaction (either the buyer or the seller) possesses more information than the other party.
- Evaluate possible government responses, including legislation, regulation and provision of information.

Asymmetric information exists when the one party in an economic transaction (either the buyer or the seller) knows important information about the good or service that he withholds from the other party.

- Without perfect knowledge, buyers may not buy the optimal quantity of a product, thus resources may be misallocated towards its production and consumption.
- Without all the information about a product, Demand (marginal private benefit) may be greater than what is socially optimal (marginal social benefit), resulting external costs for society caused by consumers demanding too much of certain goods.

Market Failures arising from Information Asymmetry

Adverse Selection	Typical market failure in the market for insurance; if the buyer of insurance does not share with the insurer complete information about the level of risk he or she presents, insurance will be provided at too low a cost to too many risky individuals. The cost of covering the dishonest are thereby shared by the more honest customers, for whom the cost of insurance is, as a result, higher than it would be otherwise.
Moral Hazard	Also a type of information asymmetry, if the consequences of one's actions are born by society as a whole or by a third party, rather than by the individual himself, he is more likely to take risky actions that he would not take if the consequences were fully born by himself. For example, if you have a rental car with full insurance, you are more likely to drive recklessly than in your own car, on which you have a high co-pay.

The US Financial Crisis as a Market Failure

What follows is a short interpretation of how the global financial crisis of 2007-2008 was the result of information asymmetry and therefore a market failure

- In the US and other countries, households were offered “subprime” loans, which allowed those who would not have traditionally qualified for a home loan to borrow money and buy a house.
- Borrowers were told that the debt they were taking on would not be a problem due to the fact that “home prices always rise”, information that was thought to be factual by most who bought homes at the time.
- Banks "bundled" these loans into securities that they sold to investors all over the world, who assumed that the lending banks were correct in their assumption that house prices would continue to rise.
- Developers built houses in record numbers based on the assumption that they'd be able to sell them at higher and higher prices.
- Supply of houses grew faster than demand, and eventually house prices began to fall.
- Borrowers found they could not make their monthly payments because their loans were "adjustable rate" meaning they required higher payments over time, causing foreclosures to increase and the supply of houses for sale to grow even more, forcing prices down even further.
- Now investors and banks all over the world hold securities made up of bad loans to Americans that were made based on the incorrect assumption that house prices would always rise. With bad assets on their "balance sheets" banks are unable to make new loans to consumers and firms, so spending in the economy has slowed, meaning recession and high unemployment

The asymmetric information at the root of the financial crisis was the belief that “home prices always rise”. When this turned out to be false, there were too many homes on the market and trillions of dollars in households' investments were lost, throwing the global economy into a recession.

Abuse of monopoly power (HL only)

- Explain how monopoly power can create a welfare loss and is therefore a type of

- market failure.
- Discuss possible government responses, including legislation, regulation, nationalization and trade liberalization.

The final type of market failure we will examine is the abuse of monopoly power by firms that control a large share of a particular market.

Monopoly power exists when a single firm controls a large share of the total market for a particular good, that firm is able to charge a higher price and produce a lower quantity than what is socially optimal.

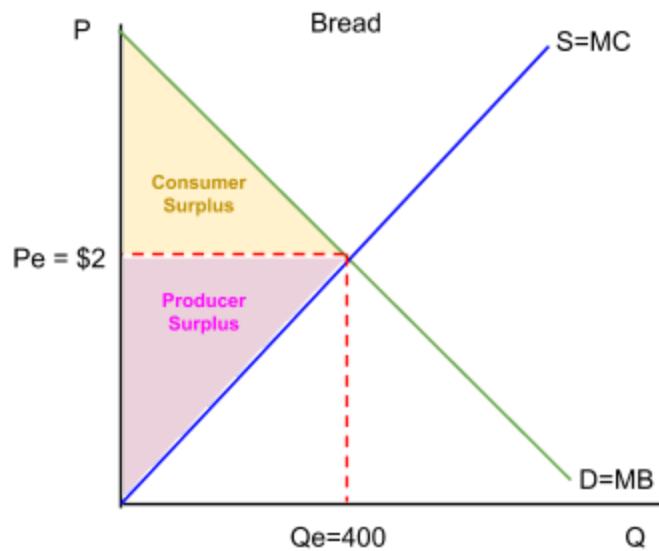
The source of monopoly power arises from a large firm's price-making abilities.

- In more competitive markets, hundreds of small firms compete with one another for the business of consumers.
- Competition forces firms to produce their goods efficiently (at a low cost) and sell their goods for a low price
- Without competition, monopolists are not forced to produce at the lowest cost, nor do they have to sell for the lowest price.

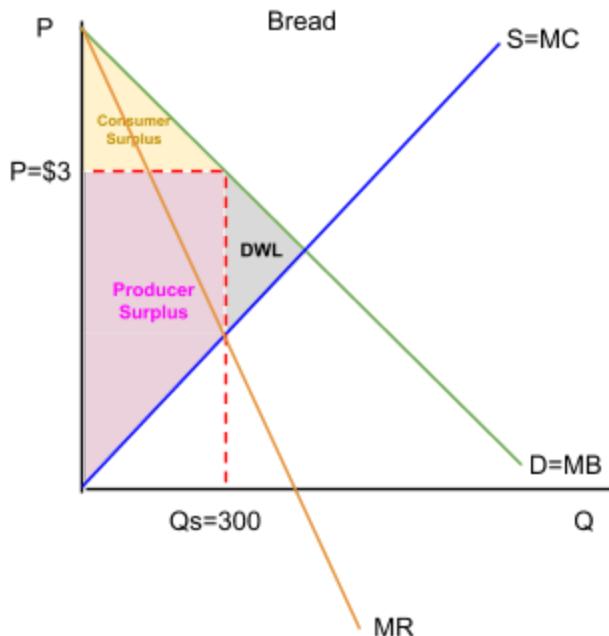
Monopolists (or firms with significant market power), are both productively and allocatively inefficient, since without competition, such firms are able to charge higher prices and produce smaller quantities!

A monopolist's price-making power allows it to produce a lower quantity and charge a higher price than what is achieved in a more competitive market.

In the graphs below we see two hypothetical markets for bread. In the first graph, hundreds of independent bakeries compete with one another to produce bread, resulting in an equilibrium at the intersection of market supply and market demand.



In the graph below a single bakery dominates the entire market, and therefore it produces a smaller quantity and charges a higher price than what is socially optimal.



300 loaves of bread is an inefficient quantity because the marginal benefit of the 300th loaf of bread is greater than the marginal cost to the bakery of producing it. Because it has monopoly power, the bakery under-produces bread and charges a higher price.

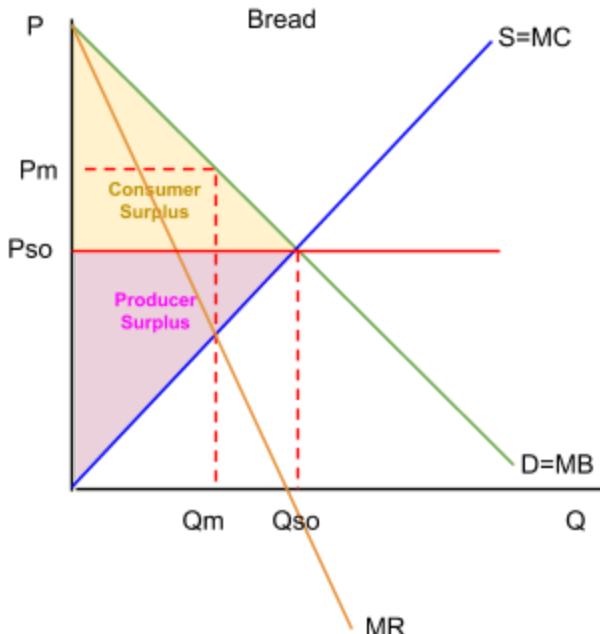
In the competitive market, the price and quantity are always determined by the intersection of demand and supply, which represent MSC and MSB, and therefore the competitive market equilibrium is allocatively efficient.

A monopolist, on the other hand, will produce at a level based on its marginal revenue and marginal cost, rather than on consumers' demand. Therefore, the monopolist will charge a higher price and produce a lower quantity than is achieved in a competitive market. Resources are under-allocated towards a monopolist's output, therefore the abuse of monopoly power is a market failure.

Government responses to monopoly power - price controls

If a government wishes to increase efficiency in an imperfectly competitive market, a price ceiling can be used to increase the quantity supplied and reduce the price.

Unlike in a perfectly competitive market, in which a price ceiling reduces quantity supplied, a price ceiling will lead a monopolistic firm to increase its output. Examine the graph for the monopolistic vegetable market below.



Effects of a price ceiling in the monopolistic market

- The red line represents a maximum price set by government below the monopolist's equilibrium price of P_e . The price ceiling is set at the socially optimal price, where $MB=MC$.
- Because the firm can no longer charge above P_{so} , the P_{so} line becomes the marginal revenue (MR) for the firm up to where it crosses the demand curve
- The profit-maximizing monopolist will produce where $MR=MC$, which now occurs at Q_{so}
- Q_{so} is where $MB=MC$, meaning this market is now producing the allocatively efficient level of output
- Deadweight loss is eliminated following the imposition of the price ceiling. The monopolist is maximizing its profits by producing at the socially optimal level of output.

Price ceilings set at or close to the socially optimal price (where $MB=MC$) can increase total surplus in a monopoly market. The firm earns less economic profit, but consumer surplus increases more than producer surplus decreases.

1.5.1 Production and Costs (HL only)

Introduction to the theory of firm behavior

Firms in a market economy are the providers of the goods and services that households demand in the product market. The incentive that drives firms to provide households with products is that of **profit maximization**.

Profit maximization: The goal of most firms is to maximize their profits. To do so, they must produce at a level of output at which the difference between their revenues and their costs is maximized.

$$\text{Economic Profit} = \text{Total Revenue (TR)} - \text{Total Cost (TC)}$$

To determine whether it will earn a profit at a particular level of output, a firm must, therefore, consider two economic variables: its costs and its revenues

Economic costs are the explicit money payments a firm makes to resources owners for the use of land, labor and capital, including

- variable costs (payments for those resources which vary with the level of output)
- and fixed costs (payments for those resources which do not vary in quantity with the level of output),

as well as the opportunity cost of the business owner (a normal profit)

$$\text{Total Cost} = \text{Variable Costs} + \text{Fixed Costs}$$

Economic Revenue is the money income a firm earns from the sale of its products to households. At a particular level of output, a firm's total revenue equals the price of its product times the quantity sold.

$$\text{Total Revenue} = \text{Price} \times \text{Quantity}$$

Short-run versus long-run costs of production

When examining a firm's costs, we must consider two periods of time.

- **The short-run:** The period of time in which firms can vary only the amount of labor and the raw materials it uses in its production. Capital and land resources are fixed, and cannot be varied.
 - Example: When the demand for American automobiles fell in the late 2000s, Ford and General Motors responded in the short-run by reducing the size of their workforces.
- **The long run:** The period of time over which firms can vary the quantities of all resources they use in production. The quantities of labor, capital and land resources can all be varied in the long run.

- Example: When demand for American automobiles remained weak for over two years, Ford and General Motors began closing factories and selling off their capital equipment to foreign car manufacturers.

A firm's **variable costs** are those that change in the short-run as the firm changes its level of output. **Fixed costs**, on the other hand, remain constant as output varies in the short-run. In the long run, all costs are variable, since all resources can be varied.

Production in the short run: the law of diminishing returns

- Distinguish between the short run and long run in the context of production.
- Define total product, average product and marginal product, and construct diagrams to show their relationship.
- Explain the law of diminishing returns.
- Calculate total, average and marginal product from a set of data and/or diagrams.

The short-run production function

The primary determinant of a firm's short-run production costs is the productivity of its short-run variable resources (primarily the labor the firm employs).

Productivity is defined as the output produced per unit of input

- The greater the average product of variable resources, the lower the average costs of production in the short-run
- The lower the productivity of the variable resource, the higher the average costs of production
- Since in the short-run, only labor and raw materials can be varied in quantity, labor is the primary variable resource

The law of diminishing marginal returns

The **law of diminishing returns** states that as more and more of a variable resource (usually labor) is added to fixed resources (capital and land), the marginal product of the variable resource will decrease. In other words, the marginal product of labor will fall as more labor is employed in the short-run (when the quantity of capital and land are fixed).

The table below presents some of the key measures of productivity we must consider when determining short-run costs.

Productivity: The amount of output attributable to a unit of input.	
Examples of productivity:	"Better training has increased the output per hour of workers" "The new robot can produce more units per minute than the previous model" "Adding fertilizer has increased the number of bushels of corn we can farm on each acre farmland"
Total product (TP)	TP is the total output of a particular firm at a particular period of time. Example of TP: "After hiring more workers the firm's total product increased."
Marginal product of labor (MP_L)	$MP_L = \Delta TP / \Delta Q_L$ The change in total product divided by the change in the quantity of labor
Average product of labor (AP_L)	$AP_L = TP / Q_L$ The output per worker, or the total product divided by the quantity of labor employed

Productivity in the short run

Assume a bakery with three ovens wishes to start making bread. To do so, it must hire workers. How many workers should the bakery hire? That depends on the productivity of the labor as more workers are added to the three ovens.

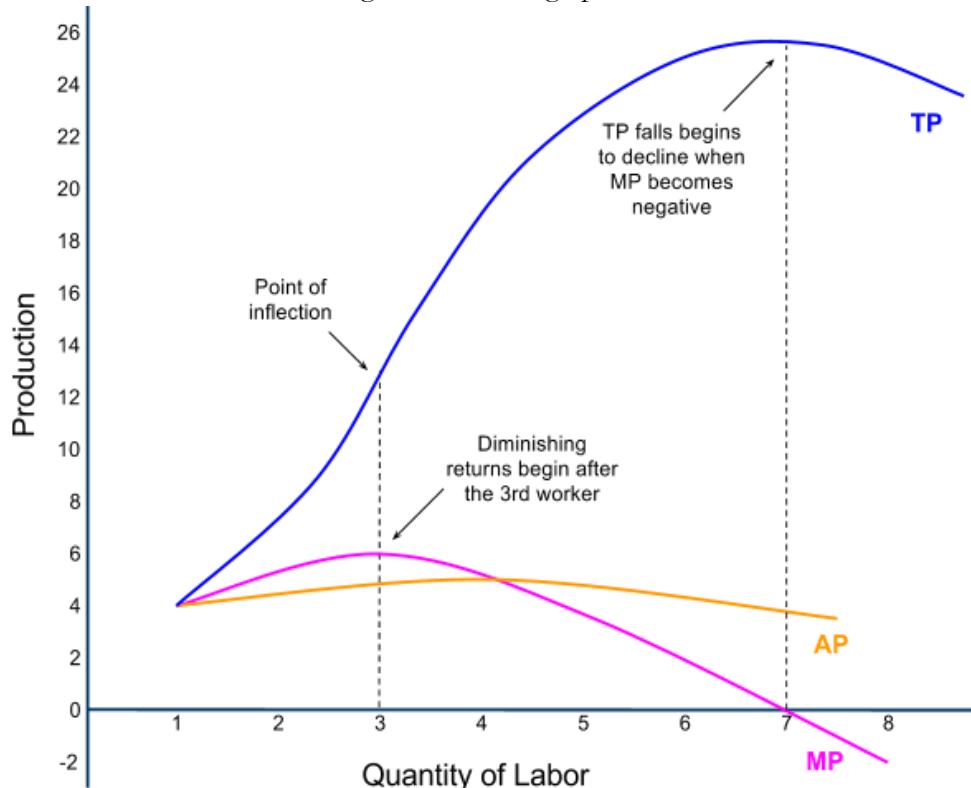
Quantity of labor (Q _L)	# of ovens	Total Product	Marginal Product (= $\Delta TP / \Delta Q_L$)	Average Product (= TP / Q_L)
0	3	0	-	-
1	3	4	4	4
2	3	9	5	4.5
3	3	15	6	5
4	3	20	5	5
5	3	24	4	4.8
6	3	26	2	4.33
7	3	26	0	3.7
8	3	24	-2	3

The table presents an example of an individual firm's short-run production. Notice the following:

- Total product increases as more workers are hired, until the 6th worker, beyond which total product remains flat and begins decreasing with the 8th worker hired.

- Marginal product (the output contributed by the last worker hired) increases until the 4th worker and then marginal product begins decreasing.
- Average product (the output per worker) increases until the marginal product becomes lower than AP (at the 5th worker) and then begins decreasing.
- The productivity of labor is at its greatest at around 3 or 4 workers, which means the bakery's average costs will be minimized when employing approximately 4 workers.

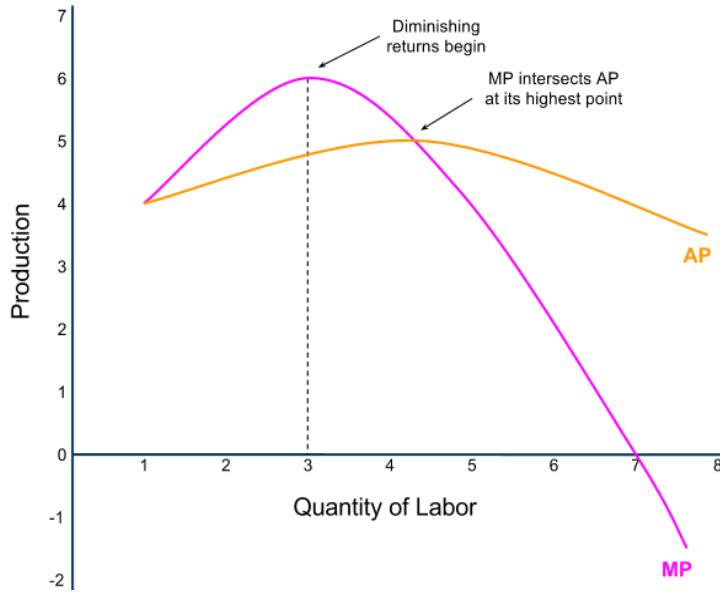
The data from our productivity table can be plotted in a graph, with the quantity of labor on the horizontal axis and the total, marginal and average product on the vertical axis.



Key observations about short-run production relationships:

- The MP is the rate of change in the TP. As MP is increasing, TP becomes steeper, but when MP decreases, TP becomes flatter. When MP becomes negative, TP begins decreasing.
- MP intersects AP at its highest point. Whenever $MP > AP$, AP is increasing, but when $MP < AP$, AP is decreasing.
- The bakery begins experiencing diminishing marginal returns (the output of additional workers begins decreasing) after the third worker

If we look more closely at just the marginal product and average product curves, we can learn more about the relationship between these two production variables.



Observe the following about the marginal product and the average product of labor in our bakery:

- With only three ovens in the bakery, the output attributable to the 4th – 8th worker becomes less and less.
- This is because there is not enough capital to allow additional workers to continue to be more and more productive!
- Up to the 5th worker, adding additional workers causes the average product to rise, since the marginal product is greater than the average.
- But beyond the 5th worker, diminishing returns causes marginal product to fall at such a rate that it pulls average output down with it. Worker productivity declines rapidly after four workers.
- A bakery wanting to minimize costs will not hire more than four workers, since the average output of the workers it hires beyond the 4th decreases (meaning the bakery pays more for each loaf of bread additional workers produce).

Costs of production in the short run

- Explain the distinction between the short run and the long run, with reference to fixed factors and variable factors.
- Distinguish between total costs, marginal costs and average costs.
- Draw diagrams illustrating the relationship between marginal costs and average costs, and explain the connection with production in the short run.
- Explain the relationship between the product curves (average product and marginal product) and the cost curves (average variable cost and marginal cost), with reference to the law of diminishing returns.
- Calculate total fixed costs, total variable costs, total costs, average fixed costs, average variable costs, average total costs and marginal costs from a set of data and/or diagrams.

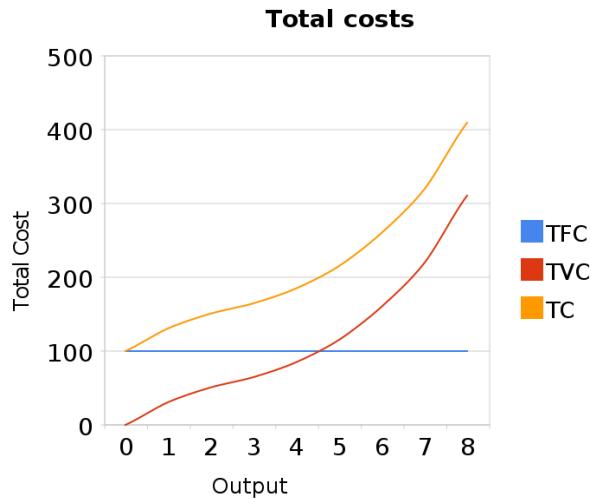
A firm faces both fixed and variable costs in the short run. The table below presents some of the costs a firm faces, and indicates whether they are fixed costs or variable costs.

Resource costs in the short-run	
Fixed Costs in the short-run	Rent - the payment for land: Rent is fixed in the short-run since firms cannot add this resource to production. Rents must be paid regardless of the level of the firm's output.
	Interest - the payment for capital: Interest is fixed in the short-run since firms cannot add this resource to production. Interest must be paid on loans regardless of the level of the firm's output.
	Normal profit - the minimum level of profit needed just to keep an entrepreneur operating in his current market. If he does not earn normal profit, an entrepreneur will direct his skills towards another market. Normal profit is a cost because if a firm does not earn normal profit, it is not covering its costs and may shut down.
Variable Costs in the short-run	Wages - the payment for labor: Wages are variable in the short-run, since firms can hire or fire workers to use existing land and capital resources. Wage costs increase when new workers are hired, and decrease when workers are laid off.
	Transportation costs: Firms pay lower transport costs at lower levels of output.
	Raw material costs: vary with the level of output
	Manufactured inputs: fewer parts are needed from suppliers when a firm lowers output.

A firm's total costs include the costs of labor (variable costs) and the costs of capital and land resources (fixed costs).

Total Costs in the Short-run
Total fixed costs (TFC): These are the costs a firm faces that do not vary with changes in short-run output. Could include rent on factory space, interest on capital (already acquired).
Total variable costs (TVC): These are the costs a firm faces which change with the level of output in the short-run. Could include payment for raw materials, fuel, power, transportation services, wages for workers, etc....
Total cost: TFC + TVC at each level of output

The graph below shows a hypothetical firm's short-run total costs of production. The shapes of the curves reflect the law of diminishing marginal returns.



From the graph above we know:

- The firm pays total fixed costs of \$100 (this may include its rent and the interest it owes the bank for the capital it borrowed when starting the business)
- When producing 0 units of output, the firm's total cost is \$100, and it's total variable cost (the cost of any labor it employs) is \$0
- As the firm begins producing output, it must hire workers, so its total variable cost increases.
- At first TVC increases at a decreasing rate as the first few workers hired are more and more productive, but then...
- TVC increases at an increasing rate as additional workers become less and less productive due to diminishing marginal returns
- The firm's total cost (TC) is the sum of its fixed costs and its variable costs.
- Total costs (TC) always increase as output increases since more and more inputs (labor, in this case) are needed to continue producing more output.

Per-unit costs in the short-run

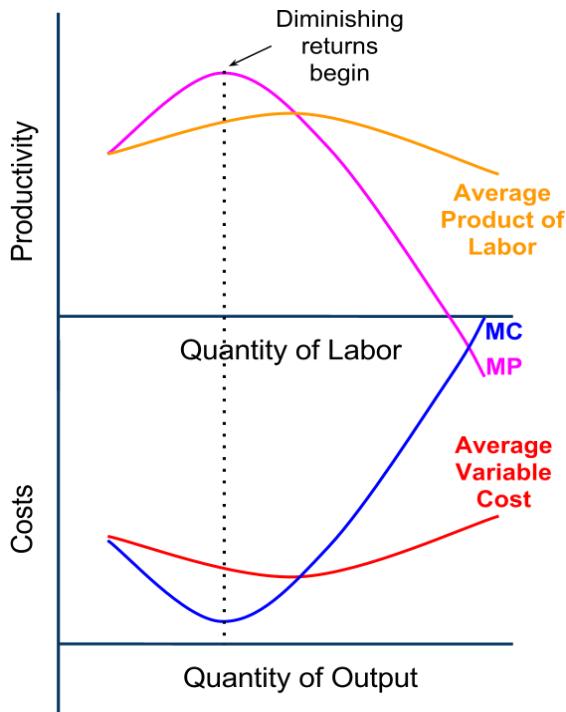
A firm makes decisions about its level of output based not on total costs, rather on per-unit and marginal costs.

Per-unit Costs in the Short-run	
Average fixed cost: $AFC = TFC \div Q$	AFC will decline as output rises, but it will never increase. This is because the fixed cost (which never goes up) is “spread out” as output increases. This is called “spreading the overhead”
Average variable cost: $AVC = TVC \div Q$	For simplicity, we will assume that labor is the only variable input, the labor cost per unit of output is the AVC
Average total cost: $ATC = TC \div Q$	Sometimes called unit cost or per unit cost. ATC also equals AFC + AVC

Marginal Cost: $MC = \Delta TVC \div \Delta Q$	The additional cost of producing one more unit of output.
--	---

From short-run productivity to short-run costs

As worker productivity increases, firms get "more for their money," meaning per-unit and marginal cost decrease. When productivity decreases, costs increase. The graph below shows the general relationship between the productivity of labor and a firm's short-run costs of production.



Observe from the graphs above that:

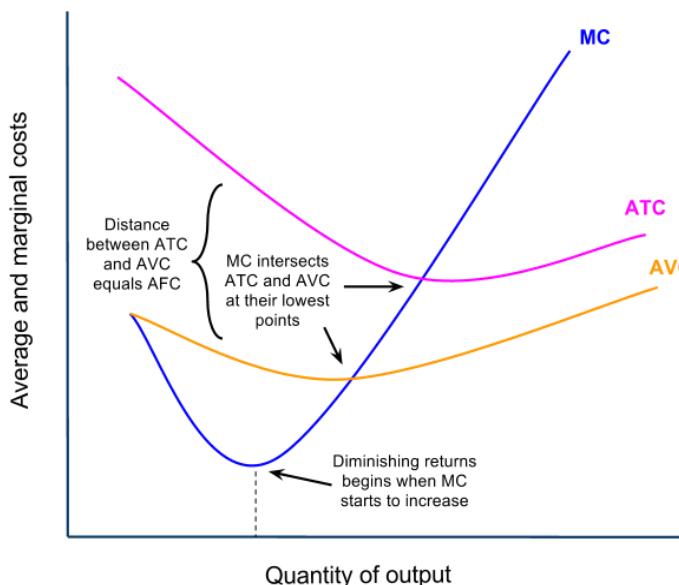
- When worker productivity is rising the firm's per unit costs are falling, since they're getting more output for each dollar spent on worker wages.
- When marginal product is increasing (increasing returns) marginal cost is falling.
- When average product is rising, average variable cost is falling.
- When MP and AP are maximized, MC and AVC are minimized
- When workers begin experiencing diminishing returns, MP falls and MC begins to rise.
- MP intersects average product at its highest point, and MC intersects average variable cost at its lowest point

Graphing per-unit costs in the short-run

The law of diminishing marginal returns dictates the relationships between a firm's short-run per unit costs

Short-run Cost Relationships	
ATC=AFC + AVC	The vertical distance between ATC and AVC equals the AFC at each level of output.
MC and ATC/AVC	MC intersects both AVC and ATC at their minimum. This is because if the last unit produced cost less than the average, then the average must be falling, and vice versa (just like your test scores!)
MC and diminishing returns	MC is at its minimum when MP is at its maximum, because beyond that point diminishing returns sets in and the firm starts getting less for its money!

The graph below shows an individual firm's short-run marginal cost (MC), average variable cost (AVC) and average total cost (ATC).



Shifts in the short-run costs of production

Anything that causes a firm's variable or fixed costs to change can cause a shift in the firm's short-run costs of production.

If workers become more productive, MC, AVC and ATC will all shift down, meaning a firm can produce additional units at a lower cost and its per unit cost will also be lower.

A change in any of the firm's resource costs will shift its short-run cost curves.

- A decrease in the wage rate will shift MC, AVC and ATC down.
- An increase in the wage rate will shift MC, AVC and ATC up.
- A change in raw material prices
- A per unit tax will shift cost curves up
- A per unit subsidy will shift cost curves down
- Higher rent will shift ATC up, but will not affect AVC or MC (since rent is a fixed cost, not a variable cost)

Costs of production in the long run

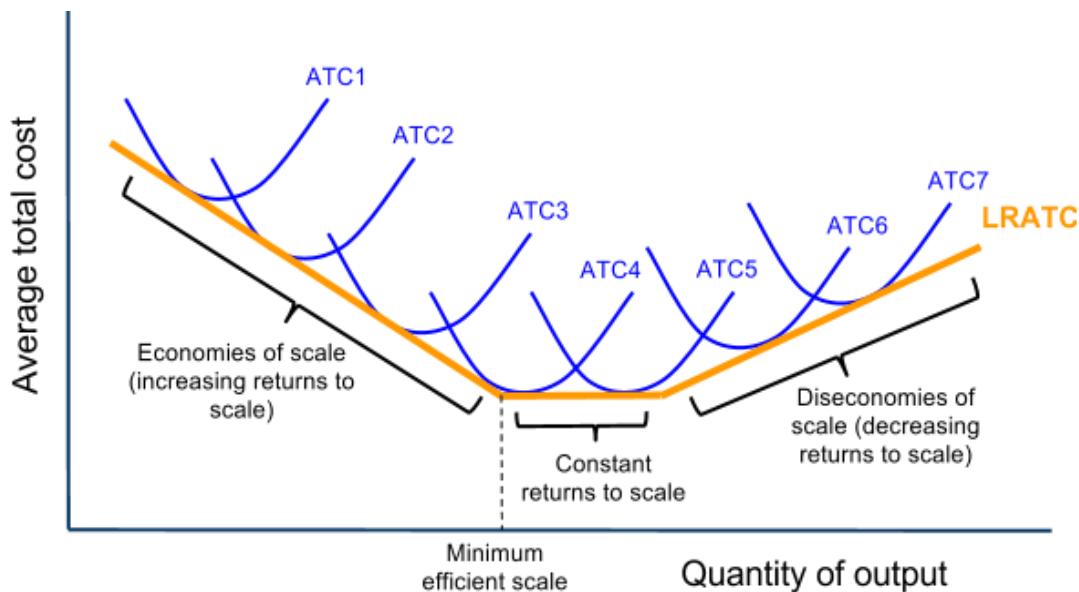
- Distinguish between increasing returns to scale, decreasing returns to scale and constant returns to scale.
- Explain the relationship between short-run average costs and long-run average costs.
- Explain, using a diagram, the reason for the shape of the long-run average total cost curve.
- Explain factors giving rise to economies of scale, including specialization, efficiency, marketing and indivisibilities.
- Explain factors giving rise to diseconomies of scale, including problems of coordination and communication.

In the **long run** all resources are variable, including capital and land. Firms can open new factories, acquire more capital, and expand into new land, or they can shut down factories, downsize their land, and sell off unused capital.

Because capital and land are variable in the long run, the law of diminishing returns no longer applies. A firm's long-run average costs depend on how productivity of land, labor and capital change as a firm's output increases in the variable plant-period. As a firm's output increases in the long run, the firm's ATC will initially decrease, but eventually increase, based on the following concepts.

Three ranges of a firm's long-run Average Total Cost curve
Increasing Returns to Scale (Economies of scale): When a firm receives increasing amounts of output per additional unit of inputs (land, labor and capital). Average costs decrease as output increases. Explanations: better specialization, division of labor, bulk buying, lower interest on loans, lower per unit transport costs, larger and more efficient machines, etc...
Constant Returns to Scale: When a firm receives the same amount of output per additional unit of input. Average costs do not change as output increases.
Decreasing Returns to Scale (Diseconomies of Scale): When a firm becomes "too big for its own good". The output per unit of input decreases as more inputs are added. Average costs increase as output increases. Explanations: Control and communications problems, trying to coordinate production across a wide geographic may make firm less efficient.

The concept of **economies of scale** explains why a firm adding new plants and capital equipment to its production will become more efficient as it expands.



Examine the long-run average total cost curve in the graph:

- The small curves represent the short-run ATC curves experienced as the firm opens several factories in the long run (from 1 to 7 factories).
- As the firm opens its first 4 factories, its ATC continuously decreases. The firm is becoming more efficient in its production.
- With the 5th factory, the firm is no longer experiencing increasing returns, and instead has experienced constant returns to scale.
- With the 6th and 7th factories, the firm's ATC is rising, indicating it is becoming less efficient. The firm is experiencing decreasing returns to scale.

As a firm increases its output in the long run (the variable plant period), it at first becomes more and more efficient, but eventually inefficiencies cause the ATC to rise, as the firm gets “too big for its own good.”

Economies of scale / Increasing returns to scale

When a firm increases its output in the long run it is adding new capital and land. If increasing these inputs leads to a proportionally larger increase in output, then the firm is experiencing **economies of scale**, or **increasing returns to scale**. Another way to look at this is if a firm doubles its inputs and its output more than doubles, then the firm experiences economies of scale.

Reasons for economies of scale include:

- Better prices for raw materials such as plastic and rubber parts for the toys due to larger bulk orders made by the firm as it grows.
- Lower costs due to higher quality and more technologically advanced machinery operating in larger factories.
- Lower average shipping and transportation costs as the firm produces and ships

- larger quantities of toys to the market when operating four factories than when operating only one.
- More favorable interest rates from banks for new capital as the firm becomes larger and therefore more "credit worthy".
- More bargaining power with labor unions for lower wages as the firm employs larger numbers of factory workers.
- Improved manufacturing techniques and more highly specialized labor, capital and managerial expertise.

Diseconomies of scale / Decreasing returns to scale

When an increase in inputs results in a proportionally smaller increase in output (for example, a doubling of inputs leads to less than a doubling of output), a firm is experiencing **diseconomies** of scale, also called **decreasing returns to scale**.

Reasons for diseconomies of scale include:

- Management inefficiencies
- Communication inefficiencies
- Corruption / graft
- Geographic, cultural or language inefficiencies of a company being spread out across international borders

Minimum efficient scale

A firm's **minimum efficient scale** is the level of output at which its long-run average total costs stop decreasing and become constant or start increasing. It is the minimum size a firm must achieve to produce its output at the lowest possible cost.

For example, in the retail market there are many firms that are big, but not big enough to compete with Walmart, America's largest retailer. Because their costs are higher than Walmarts, they are unable to compete on price. We can say that such firms have not achieved minimum efficient scale, which Walmart achieved long ago. Therefore, they are unable to achieve the low costs that Walmart is able to.

Total revenue, average revenue and marginal revenue

- Distinguish between total revenue, average revenue and marginal revenue.
- Draw diagrams illustrating the relationship between total revenue, average revenue and marginal revenue.
- Calculate total revenue, average revenue and marginal revenue from a set of data and/or diagrams.

Costs are only half the calculation a firm must make when determining its level of economic profits. A firm must also consider its revenues.

Revenues are the income the firm earns from the sale of its good.

- Total Revenue** = the price the good is selling for X the quantity sold
- Average Revenue** = The firm's total revenue divided by the quantity sold, or simply

- the price of the good
- **Marginal Revenue** = the change in total revenue resulting from an increase in output of one unit

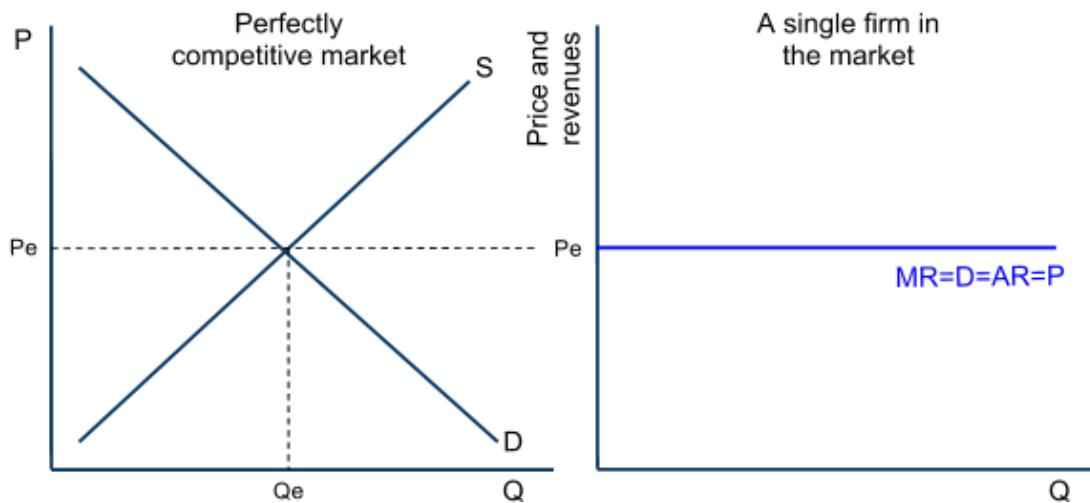
Depending on the market structure in which a firm produces its product, it is either a price-taker or a price-maker.

- For some firms, the price it can sell additional units of output for never changes. These firms are known as “price-takers”, and sell their output in highly competitive markets
- For other firms, the price must be lowered to sell additional units of output. These firms are known as “price-makers” and have significant market power, selling their products in markets with less competition.

Revenues for a perfect competitor

A firm selling its product in a perfectly competitive market is a “**price-taker**”. This means the firm can sell as much output as it wants at the equilibrium price determined by the market.

- The marginal revenue the firm faces, therefore, is equal to the price determined in the market.
- The average revenue is also the price in the market.
- The $MR=AR=P$ line for a perfectly competitive firm also represents the demand for the individual firm’s product. Because a perfectly competitive seller is one of hundreds of firms selling an identical product, the firm cannot raise its price above that determined by the market.



Demand for the perfectly competitive firm’s output:

- Demand is perfectly elastic
- The firm has no price-making power.
- The price in the market equals the firm’s marginal revenue and average revenue

Revenues for an imperfect competitor

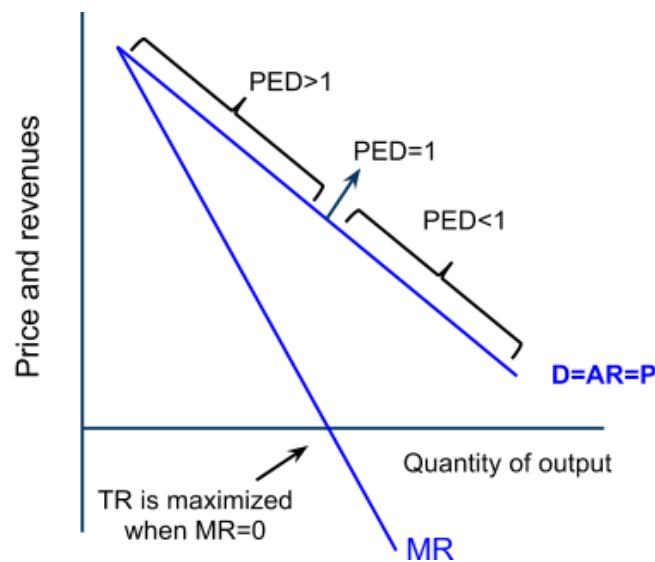
A firm with a large share of the total sales in a particular market is a “price-maker”, because to sell more output, it must lower its prices. For this reason...

- The marginal revenue the firm faces is less than the price at each level of output.
- The average revenue is also the price in the market.
- Because a firm with market power is selling a unique or differentiated product, it faces a downward sloping demand curve.

Consider the data in the table below, which shows the price, total revenue and marginal revenue for an imperfectly competitive firm:

Quantity of output (Q)	Price (P) = Average Revenue (AR)	Total Revenue (TR)	Marginal Revenue (MR)
0	450	0	-
1	400	400	400
2	350	700	300
3	300	900	200
4	250	1000	100
5	200	1000	0
6	150	900	-100
7	100	700	-200
8	50	400	-300

The firm whose revenues are depicted above will see the following Demand, Average Revenue, Marginal Revenue and Total Revenue:



Points to notice about the imperfect competitor:

- To sell more output, this firm must lower its price
- As it sells more output, its MR falls faster than the price, so the MR curve is always below the Demand curve (except at an output of 1 unit, when $MR=P$)
- The firm's total revenues rise as its output increases, until the 6th unit, when the firm's MR has become negative. MR is the change in TR, so when MR is negative, TR begins to fall.
- The firm would never want to sell more than 5 units. This would cause the firm's costs to rise while its revenues fall, meaning the firm's profits would be shrinking.
- The demand curve has an elastic range and an inelastic range, based on the total revenue test of elasticity

Economic profit and normal profit

- Describe economic profit (abnormal profit) as the case where total revenue exceeds economic cost.
- Explain the concept of normal profit (zero economic profit) as the amount of revenue needed to cover the costs of employing self-owned resources (implicit costs, including entrepreneurship) or the amount of revenue needed to just keep the firm in business.
- Explain that economic profit (abnormal profit) is profit over and above normal profit (zero economic profit), and that the firm earns normal profit when economic profit (abnormal profit) is zero.
- Explain why a firm will continue to operate even when it earns zero economic profit (abnormal profit).
- Explain the meaning of loss as negative economic profit arising when total revenue is less than total cost.
- Calculate different profit levels from a set of data and/or diagrams.

Accounting profit vs. economic profit

A firm is earning an **accounting profit** as long as its total revenues exceed its total explicit costs. For example, assume a bakery faces the following monthly costs.

Cost	Amount
Worker wages	\$4,000
Rent for retail space	\$2,000
Utilities	\$1,000
Raw materials	\$1,000
Interest on bank loan	\$500
Total explicit costs	\$8,500

Total revenues	\$9,500
----------------	---------

Assume the firm above earns \$9,500 revenue each month. The firm is earning an accounting profit of :

$$\begin{aligned} \text{Accounting profit} &= \text{Total revenue} - \text{Total explicit costs} \\ &= \$9,500 - \$8,500 = \$1,000 \end{aligned}$$

So you may say that any firm that earns an accounting profit is profitable. However, economic profit includes not just explicit, monetary costs, but also the entrepreneur's opportunity cost. In other words, what is the minimal amount of profit beyond all its explicit costs a business owner expects to make? This "cost" is called the entrepreneur's "**normal profit**".

Consider the table again, this time with the entrepreneur's normal profit and total economic costs included.

Cost	Amount
Worker wages	\$4,000
Rent for retail space	\$2,000
Utilities	\$1,000
Raw materials	\$1,000
Interest on bank loan	\$500
Total explicit costs	\$8,500
Normal profit	\$1,500
Total economic costs	\$10,000
Total revenues	\$9,500

From this table we can see that the bakery's revenues of \$9,500 are less than its total economic costs of \$10,000. This firm's **economic profit** can be calculated:

$$\begin{aligned} \text{Economic profit} &= \text{Total revenues} - \text{Total economic costs} \\ &= \$9,500 - \$10,000 = \textcolor{red}{-\$500} \end{aligned}$$

It turns out that with revenues of \$9,500 this bakery is earning economic losses of \$500. Despite all of its explicit costs being covered, the owner of the bakery does not feel it is worth her effort to continue operating the business. Unless the losses are eliminated through

higher demand or lower costs, the bakery owner will eventually shut the business down because it is not earning her an economic profit.

Profit maximization

- Explain the goal of profit maximization where the difference between total revenue and total cost is maximized or where marginal revenue equals marginal cost.

Considering its costs and revenues, a firm must decide how much output it should produce to maximize its economic profits.

- Economic Profits = Total Revenues – Total Costs
- Per-unit Profit = Average Revenue (or price) – Average Total Cost

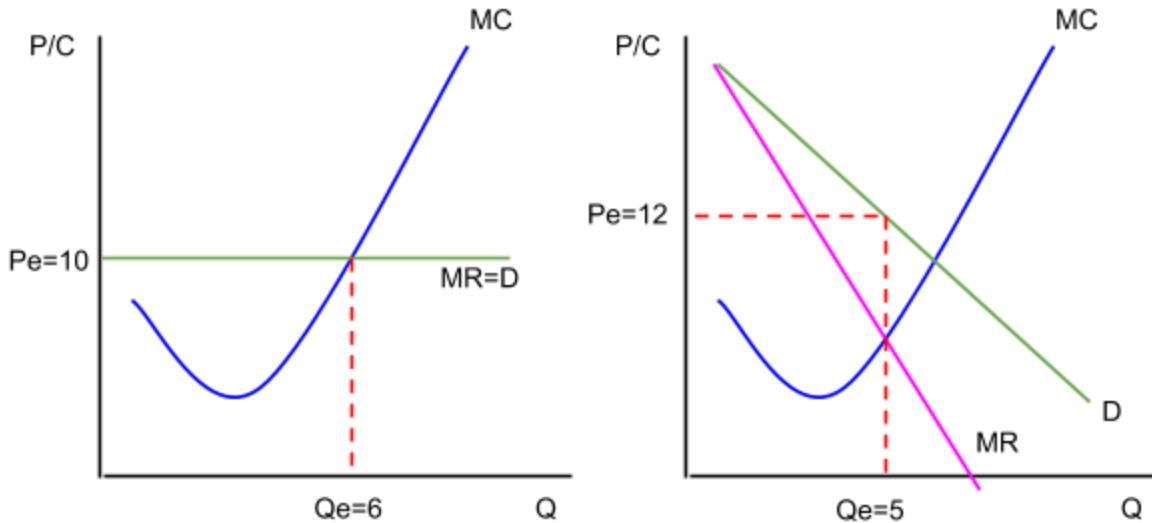
The **profit-maximizing rule** says that to maximize its total economic profits, a firm should produce at the level of output at which its marginal revenue equals its marginal cost

- For a perfect competitor, the $P=MR$, so the profit-maximizing firm should produce where its $MC=P$.
- For an imperfect competitor, the $P>MR$, so the profit-maximizing firm will produce at a quantity where its $MC=MR$.

If a firm is producing at a point where its $MR>MC$, the firm's total profits will rise if it continues to increase its output, since the additional revenue earned will exceed the additional costs. If a firm is producing at a point where $MC>MR$, the firm should reduce its output because the additional costs of the last units exceed the additional revenue. When $MC=MR$, the firm's total profits are maximized

Consider the firms below:

- The competitive, price-taking firm on the left will produce where its $MC=MR$, at a quantity of 6 units and at the market price of \$10.
- The firm on the right, a price-making monopolist, will produce 5 units and charge a higher price of \$12.



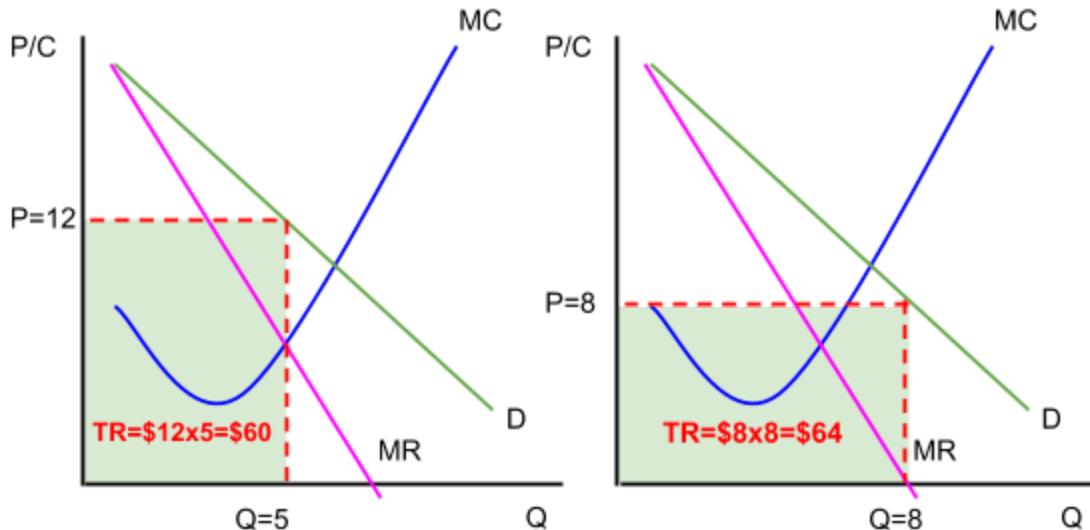
Alternative goals of firms

- Explain alternative goals of firms, including revenue maximization, growth maximization, sacrificing and corporate social responsibility

While it is typically assumed that firms are profit maximizers, there are alternative goals that firms may pursue that require them to sacrifice some level of profit towards other ends.

Revenue maximization: A firm can maximize its revenues by producing at the quantity at which the marginal revenue (MR, or the amount by which the firm's total revenues increased with the sale of the last unit produced) equals zero. At this level of output, the marginal cost will be higher than the marginal revenue (meaning the firm's profits are smaller than they could be), but the firm's level of output will be higher, perhaps allowing it to control a larger share of its market.

The firm on the left below is producing at its profit-maximizing level of output, where $MR=MC$, and earning total revenues of \$60. The firm on the right, on the other hand, is producing where $MR=0$, selling a higher quantity of output (8 units), and charging a lower price. However, its revenues are greater than if it were maximizing profits. The firm's profits will be smaller, since at 8 units the $MC>MR$, meaning the last three units produced have cost the firm more than they increased its revenues.

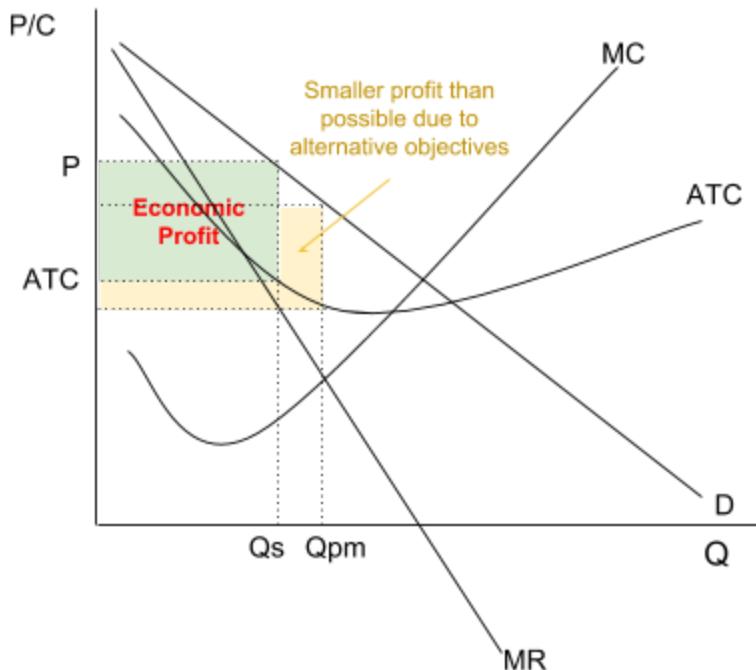


The revenue maximizing firm:

- Produces where the last unit sold earned the firm no additional revenue (where $MR=0$)
- At every level of output up to this point increases in output caused revenues to rise ($MR>0$)
- At every level of output beyond this point increases in output caused revenues to decrease ($MR<0$)
- Total revenues are maximized where $MR=0$, but the level of economic profit is smaller than it would be at a lower quantity and higher price.

Satisficing behavior: As we just learned, profit maximization is not always the main objective firms pursue. Sometimes firms may sacrifice economic profit for alternative objectives, such as corporate social responsibility, environmental sustainability and public service ends.

When a firm aims to earn a profit but not necessarily maximize its profit due to its promotion of alternative objectives, the term used is “satisficing behavior”. While there is no single “satisficing rule”, a firm intentionally earning, but not maximizing profits, could be said to be satisficing.



Observe from the graph:

- This firm is producing a smaller quantity and charging a price than if it were to produce at its profit-maximizing level
- The firm is still earning a profit, just not maximizing its profit.
- The firm may be satisfying e.g. earning enough to keep its owners (or shareholders) satisfied while pursuing other objectives
- The firm may be investing in renewable production methods (which limits output and increase cost), donating a share of its revenues to charities, sourcing its raw materials from fair trade or certified organic suppliers, etc...
- Any of these alternative objectives would result in a lower level of output, higher ATC and higher price, but may be considered satisfying behavior, since the firm is still earning a profit, just not the maximum amount possible.

1.5.2 Perfect Competition

Assumptions of the Perfectly Competitive Market Model

- Describe, using examples, the assumed characteristics of perfect competition: a large number of firms; a homogeneous product; freedom of entry and exit; perfect information; perfect resource mobility.

Introduction to market structures

Product markets come in many forms. The four market structures are introduced below.

Characteristic	Perfect Competition	Monopolistic Competition	Oligopoly	Monopoly
Number of Firms	VERY large number of firms	Fairly large number of firms	A few large firms dominate an industry	Only ONE firm. The firm IS the industry
Price-making abilities of individual firms	Each firm is so small that changes in its own output do not affect market price, i.e. firms are price takers	Firms are small relative to the industry, meaning changes in one firm's output have only a slight impact on market price	A change in one firm's output has significant impact on the market price, firms are price-makers.	Changes in the firm's output cause changes in the price, i.e. the firm is a price-maker!
Type of product	Firms all produce identical products, with no differentiation	Products are slightly differentiated. Firms will advertise to try and further differentiate product. Branding! Advertising!	Products can be identical (such as oil) or differentiated (such as Macs and Dells) Firms will likely use advertising to try and differentiate their products from competitors'	Unique product, no other firm makes anything like it.
Entry barriers	Completely free entry and exit from the industry, i.e. NO barriers to entry.	Limited barriers to entry, firms can enter or leave easily	There are significant barriers to entry	Significant barriers to entry exist, preventing new firms from entering and competing with the monopolist

Assumptions of the perfectly competitive market model

A **perfectly competitive market** is one in which:

- There is a very large number of firms,

- Selling identical products to one another,
- In which there are no barriers to entry or exit
- And in which individual firms have no control over the market price.

For example, imagine you live in New York City and want to have a cheese pizza for lunch. Let's assume...

- There are hundreds of pizza shops in New York
- Every one of them has cheese pizza on their menu
- They all pay their workers minimum wage. They all buy cheese, dough and tomato sauce at the same prices
- It is cheap and easy to open a pizza shop, just as it is to shut one down if needed.

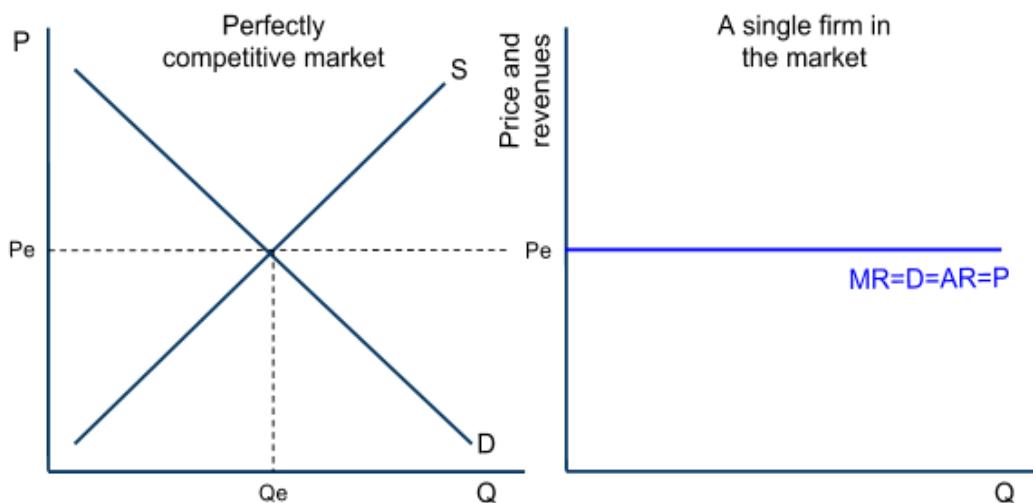
Based on these characteristics, the market for cheese pizzas in New York is close to perfectly competitive. You will pay the same price no matter where you order your pizza!

Revenue curves for the perfect competitor

- Explain, using a diagram, the shape of the perfectly competitive firm's average revenue and marginal revenue curves, indicating that the assumptions of perfect competition imply that each firm is a price taker.
- Explain, using a diagram, that the perfectly competitive firm's average revenue and marginal revenue curves are derived from market equilibrium for the industry.

Perfectly competitive firms are “price takers”. An individual firm in a perfectly competitive market has no control over the price of its own output. This is because the price is determined based on market supply and market demand.

In the graphs below shows the market supply and demand for a good on the left and the demand (D), marginal revenue (MR), average revenue (AR), and the price (P) as seen by an individual producer of the good.



Notice from the graphs:

- The demand seen by the firm is determined by the price in the market.

- Price also determines the firm's marginal revenue. The firm can sell as much as it likes at the market price and always earn the same additional revenue.
- The firm has no “price-making power” because if it raises its price, it will sell no output, and
- If it lowers its price, it will not be able to cover its costs of production.
- Demand for the individual firm's output is perfectly elastic

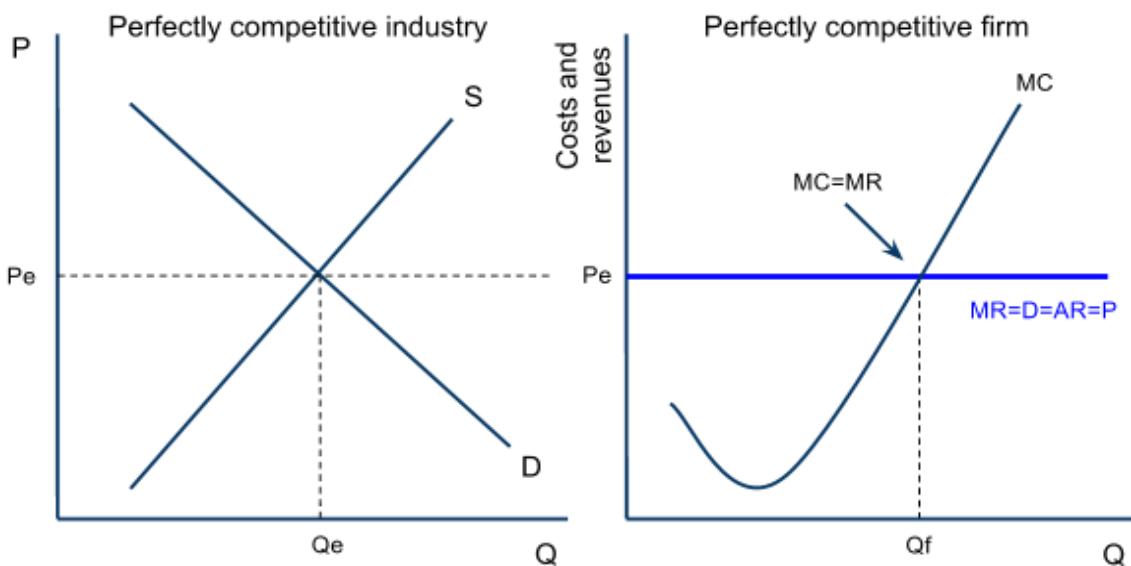
Profit maximization in the short run

- Explain, using diagrams, that it is possible for a perfectly competitive firm to make economic profit (abnormal profit), normal profit (zero economic profit) or negative economic profit in the short run based on the marginal cost and marginal revenue profit maximization rule.

In our previous unit we learned that, to maximize its economic profits at any given time, a firm should produce at the quantity at which its marginal revenue (MR) equals its marginal cost (MC)

In the graph below:

- The firm is facing marginal revenue equal to P_e , determined by the market price at any given time.
- The firm's marginal cost increases steadily due to diminishing returns (to make more pizzas in the short-run, more cooks must be hired, but because capital is fixed the marginal product of cooks decreases as more are added)
- Based on its MC and MR, the firm will maximize its profits (or minimize its losses) by producing at Q_f



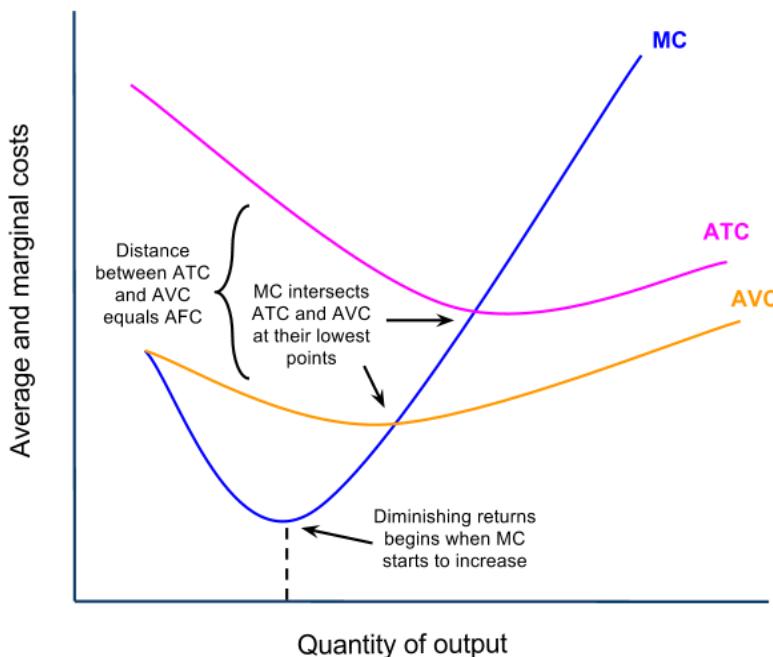
Determining short-run economic profits or loss

A firm will maximize its profits (or minimize its losses) by producing at the quantity where $MR=MC$. To determine whether a firm is actually earning profits, breaking even, or earning

losses at this quantity, we must consider both the firm's average revenue (the price) and its average total cost.

Recall from our earlier unit that a firm faces the following short-run production costs:

- Marginal Cost, which slopes upwards because of diminishing marginal returns
- Average variable cost, which is the per unit labor costs of production
- Average total cost, which is the average variable costs plus the average fixed costs (the per-unit costs of fixed capital resources)
- Recall also that MC must intersect the average cost curves at their lowest points.

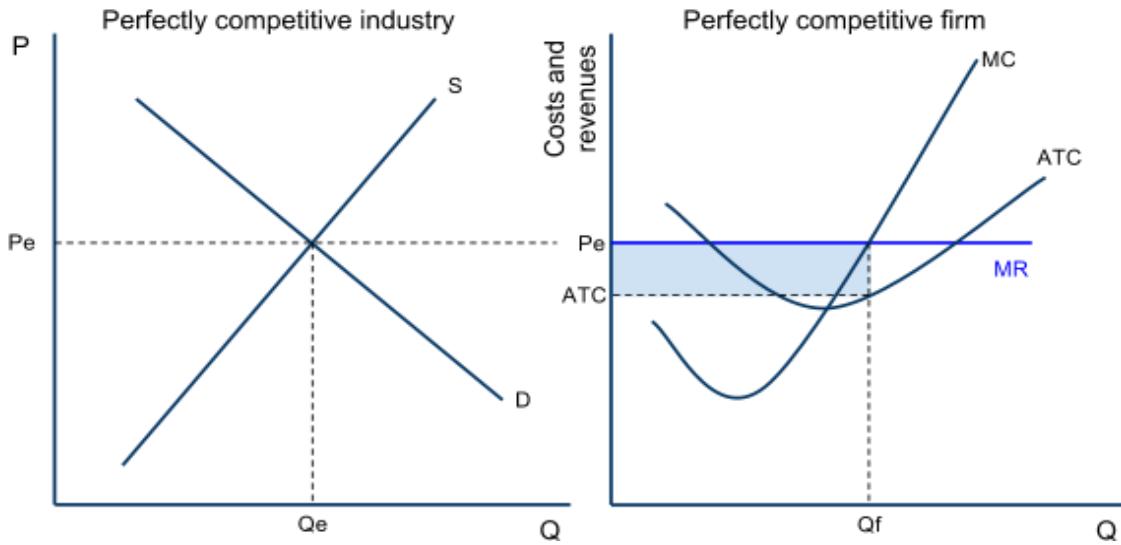


Profit maximization in the short run - the profit-earning firm

If, when producing at its $MC=MR$ point, a firm in a perfectly competitive market is selling its output for a price that is greater than its average total cost, then the firm is earning economic profits. Economic profits mean the firm is covering all of its explicit and implicit costs, and is earning additional revenue beyond these as well.

Study the graph below and note:

- The market demand is relatively high, presenting firms with a price that is greater than their ATC
- The firm's economic profits are represented by the shaded rectangle $(P-ATC)xQ$.
- The firm is maximizing its profits by producing where $MR=MC$.
- Due to the absence of entry barriers, these profits will not be sustained in the long run, as new firms will enter the market.

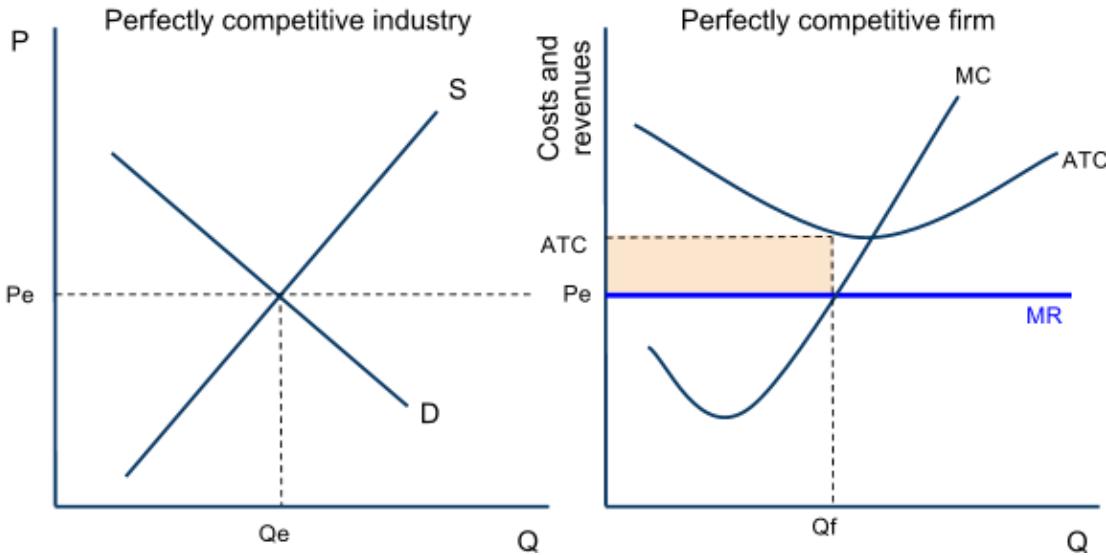


Profit maximization in the short run - the loss minimizing firm

If, when producing at its $MC=MR$ point, a firm in a perfectly competitive market is selling at a price that is lower than its average total cost, the firm will be minimizing its losses, but earning no economic profit at all. The loss-minimizing firm will either exit the industry in the long run, or hope other firms exit until the supply decreases, causing the price to rise once again.

Study the graph below and note:

- The market demand is relatively low, so the price the firm can sell its output for is below its average total cost
- The firm's economic losses are the shaded area $(ATC-P)xQ$.
- The firm is minimizing its losses by producing where $MR=MC$.
- Due to the absence of entry barriers, these losses will be eliminated in the long run as firms exit the industry to avoid further losses.

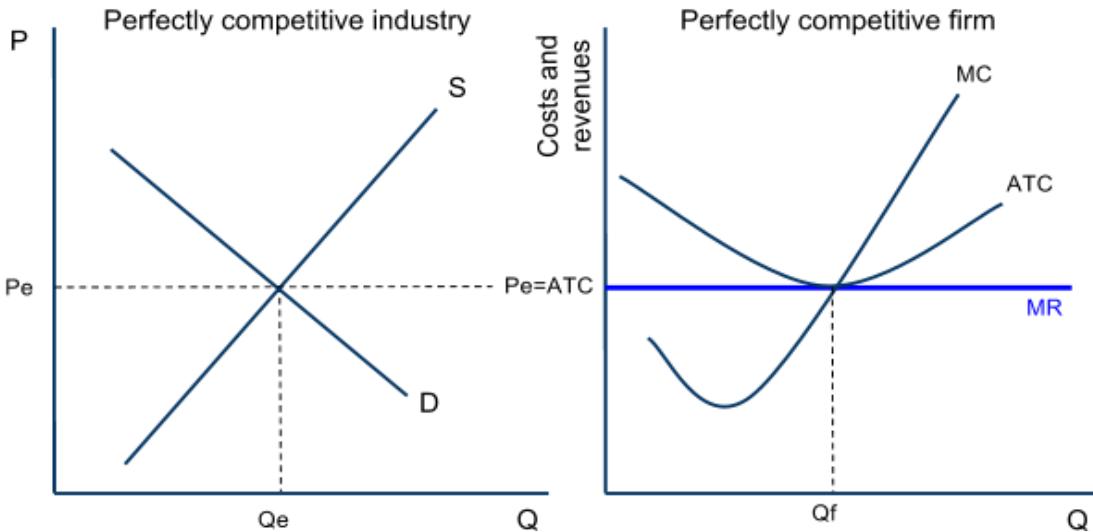


Profit maximization in the short-run - the breaking even firm

If, when producing at its $MC=MR$ level of output, the price the firm can sell its output for is exactly equal to the firm's minimum average total cost, then the best the firm can hope to do is to break even. Breaking even means a firm is covering all of its explicit and implicit costs, but earning no additional profit. The firm is earning only a normal profit

Study the graph below and note:

- The market demand and supply have set a price equal to the firm's minimum average total cost.
- The firm is just covering all its costs, meaning it is earning zero economic profits, but no losses
- If the firm produced at any quantity other than Q_f , it would earn economic losses. By producing at Q_f , it is breaking even.
- There is no incentive for firms to enter or exit this market.



Profit maximization in the long run

- Explain, using a diagram, why, in the long run, a perfectly competitive firm will make normal profit (zero economic profit).
- Explain, using a diagram, how a perfectly competitive market will move from short-run equilibrium to long-run equilibrium.

The long run is defined as the period of time over which firms can adjust their plant size in response to changes in the level of demand for their product. New firms can enter a market and existing firms can exit a market in the long run. The long run is the variable-plant period.

In perfectly competitive markets, firms can enter or exit the market in the long run.

- If economic profits are being earned, firms will be attracted to the profits and will want to enter the market
- If economic losses are being earned, some firms will wish to minimize their losses by shutting down and leaving the market
- Due to the entry and exit of firms in perfectly competitive markets, economic profits and losses will be eliminated in the long run and firms will only BREAK EVEN.

When all the firms in a perfectly competitive market are breaking even, a market is in its long-run equilibrium state. No firms will wish to enter OR exit a market in which firms are breaking even!

Profit maximization in the long run – entry eliminates profits

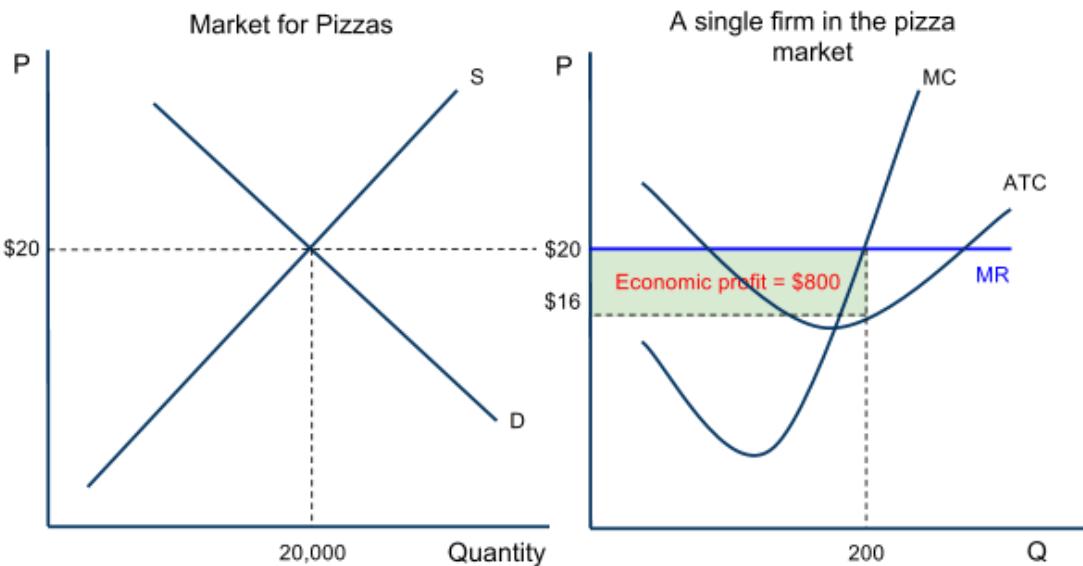
When individual firms are earning economic profits in a perfectly competitive market, new firms will be attracted to the market, leading to an increase in market supply and a fall in the price.

Consider the pizza market on the next page:

- At the current level of demand and supply, the price (\$20) per pizza is greater than

the typical firm's ATC (\$16).

- Pizza shops are making 200 pizzas each at a profit of \$4 per pizza for a total profit of \$800.
- Due to the low entry barriers, sellers of other products will be attracted to the pizza market, where easy profits can be earned.

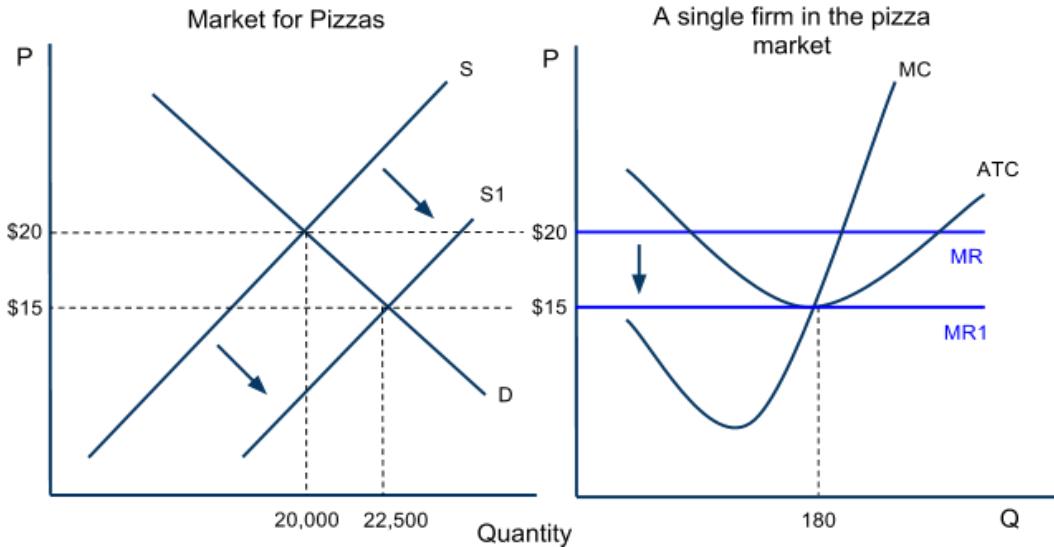


The existence of economic profits will attract new sellers to the pizza market.

- The number of sellers is a determinant of supply, so the market supply will increase
- The increased competition in the pizza market causes the price of pizzas to fall and the market quantity to increase

For the individual firm in the market:

- The price of pizza falls from \$20 to \$15.
- MR falls, causing the firm to reduce its output to maintain its $MR=MC$ level
- Economic profit is eliminated, as the price falls to the firm's minimum ATC
- The firm's output is reduced as it now faces more competition
- The market is in equilibrium again when the individual firm is only breaking even

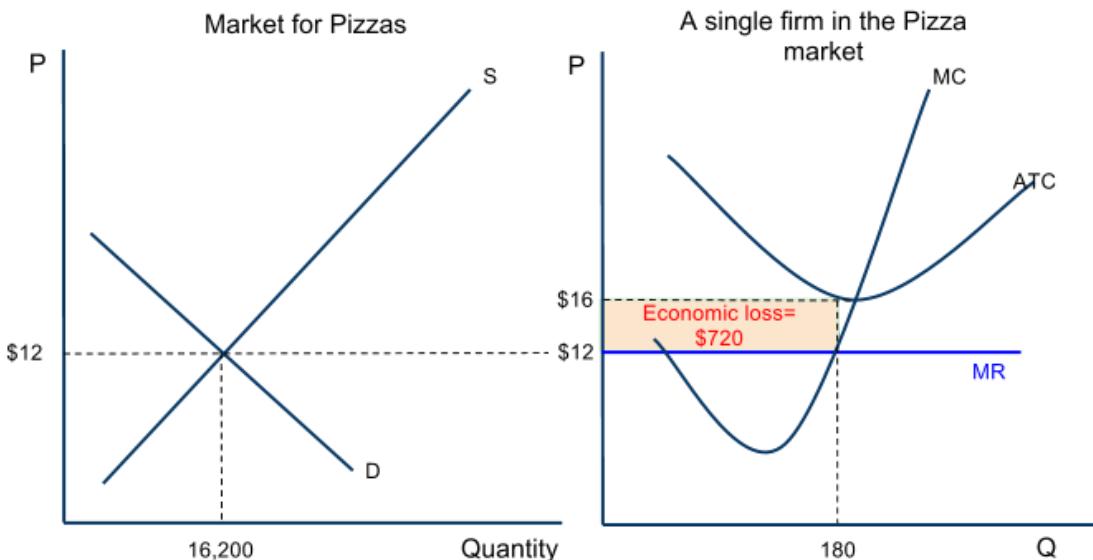


Profit maximization in the long run – exit eliminates losses

When individual firms are earning losses in a perfectly competitive market, certain firms will choose to leave the market to avoid losses and to seek profits elsewhere.

Consider the pizza market:

- At the current level of demand and supply the market price (\$12) is lower than the typical firm's ATC (\$16)
- Pizza shops are making 180 pizzas each at a loss of \$4 per pizza, for a total loss of \$720
- Due to the fact that it is easy to exit the market, some pizza shops will chose to shut down and seek profits elsewhere.

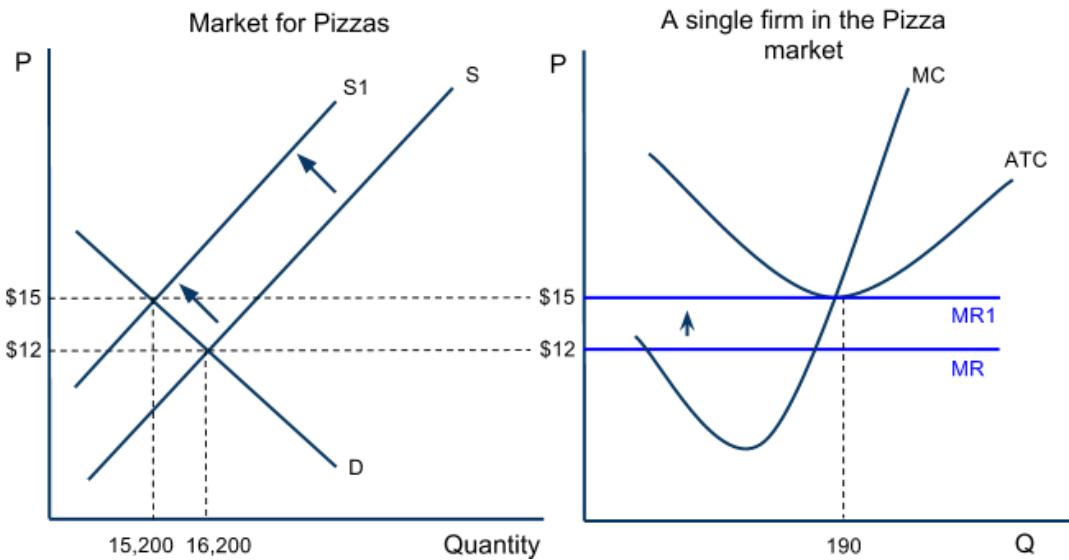


The experience of earning losses will make some firms wish to leave the market

- The number of sellers is a determinant of supply, so the market supply will decrease
- The decreased competition in the pizza market causes the price of pizzas to rise and the market quantity to decrease

For the individual firm that remains in the market:

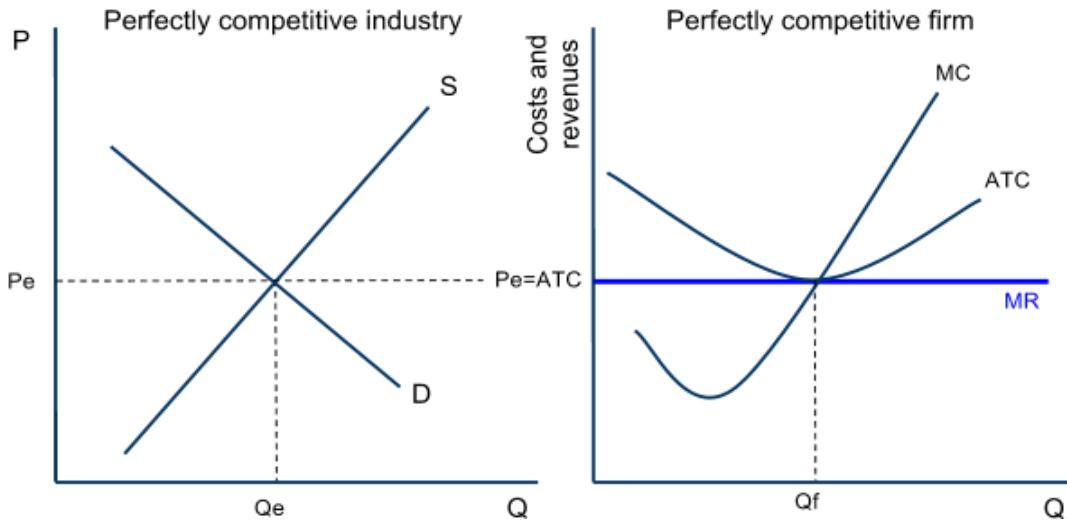
- The price of pizza rises from \$12 to \$15
- MR rises, causing the firm to increase its output to maintain its $MR=MC$ level
- Losses are eliminated, as the price rises to the firm's minimum ATC
- The firm's output increases as it now faces less competition
- The market is in equilibrium again when the individual firm is only breaking even



Long-run equilibrium in perfect competition

A perfectly competitive market is in its long-run equilibrium when the typical firm is breaking even.

- Equilibrium is defined as “a state of balance”
- If any profits or losses are being earned, a PC market is out of balance, and firms will enter or exit the market until equilibrium is restored.



Shutdown price and break-even price

- Distinguish between the short run shutdown price and the break-even price.
- Explain, using a diagram, when a loss-making firm would shut down in the short run.
- Explain, using a diagram, when a loss-making firm would shut down and exit the market in the long run.
- Calculate the short run shutdown price and the breakeven price from a set of data

So far we have said that “if losses are being earned, some firms will exit the market until the remaining firms are breaking even once again.” But this raises the question: Which firms will exit the market, and which firms will stay?

Revisiting our assumptions about perfect competition: Recall that we said that PC firms face identical costs of production. That is not 100% true, because one cost, the level of normal profit, can vary from seller to seller, even in perfect competition.

Normal profit is the implicit, subjective value of each business owner’s skills and time. Some business owners will value their efforts more highly than others, even when all the other costs faced are identical to all other business owners’ costs.

- For this reason, some sellers will be willing to tolerate greater losses for longer periods of time than other sellers.
- In other words, among firms facing identical explicit costs (wages, interests, rents), some will shut down sooner when earning losses than others due to their different levels of implicit costs (normal profit)

A firm facing economic losses has two choices:

1. Continue to operate your business, and hope that your average revenue (price) is at least high enough to cover your average variable costs (these are your operating costs in the short-run... you have to earn enough to pay your workers, at least!), OR...

- Shut down and give up your fixed costs, which are those that must be paid EVEN if you shut your business down. A firm's loss when it shuts down is its total fixed costs, those payments to owners of capital and land resources (rent for your landlord, interest owed to the bank on money you borrowed to buy capital).

These tradeoffs give business owners a clear rule for when to shut down::

- If the price of the product is lower than the firm's average variable cost, or
- If the firm's total losses when continuing to operate are greater than its total fixed costs

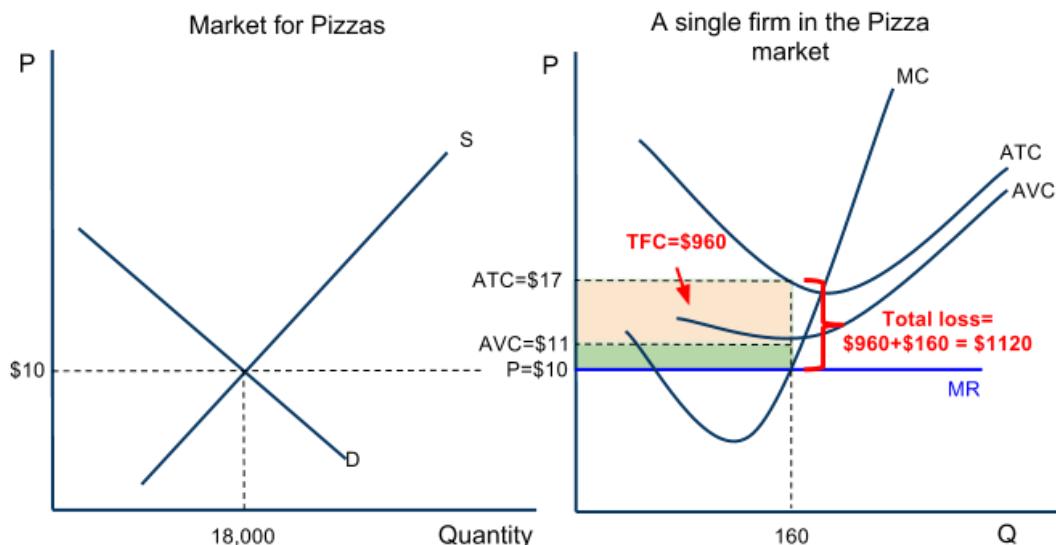
If either of these criteria are true, then a firm can always minimize its losses by shutting down and leaving the market. If neither is true, the firm should remain in the market and continue to produce, and hope that the price rises again in the future.

A firm facing losses must compare its level of losses by continuing to operate to its level of losses if it shuts down.

- Total losses if it continues to operate = $(AR-ATC) \times Q$
- Total losses if it shuts down = $(ATC-AVC) \times Q$

Consider the pizza market:

- The demand for pizzas is so low that the price (\$10) is lower than the firm's AVC (\$11). The firm cannot even afford to pay its workers.
- The firm's total losses $(17-10) \times 160$, are greater than its total fixed costs $(17-11) \times 160$. The firm would minimize its losses by shutting down
- This firm should exit the market



Efficiency in perfectly competitive markets

- Explain the meaning of the term allocative efficiency.
- Explain that the condition for allocative efficiency is $P = MC$ (or, with externalities,

$MSB = MSC$).

- Explain, using a diagram, why a perfectly competitive market leads to allocative efficiency in both the short run and the long run.
- Explain the meaning of the term productive/technical efficiency.
- Explain that the condition for productive efficiency is that production takes place at minimum average total cost.
- Explain, using a diagram, why a perfectly competitive firm will be productively efficient in the long run, though not necessarily in the short run.

In long-run equilibrium, purely competitive firms will produce the efficient level of output and price. Efficiency in economics is measured in two ways: Firms can be productively efficient and an industry can be allocatively efficient.

Productive efficiency is achieved if firms produce at their minimum average total cost. Firms are using resources to their maximum efficiency by producing their output at the lowest possible average total cost. Competition forces firms to use resources as efficiently as possible.

Allocative efficiency is achieved if a market produces at the quantity where marginal benefit equals marginal cost (where Price = Marginal Cost). The right amount of output is being produced. There is neither under or over-allocation of resources towards a good in a purely competitive industry.

- If the price were higher than the marginal cost, this is a signal that marginal benefit exceeds marginal cost and more output is desired,
- If price were lower than marginal cost, the signal from buyers to sellers is that marginal cost exceeds marginal benefit and less output is desired.
- Only when $P = MC$ is the right amount of output being produced.

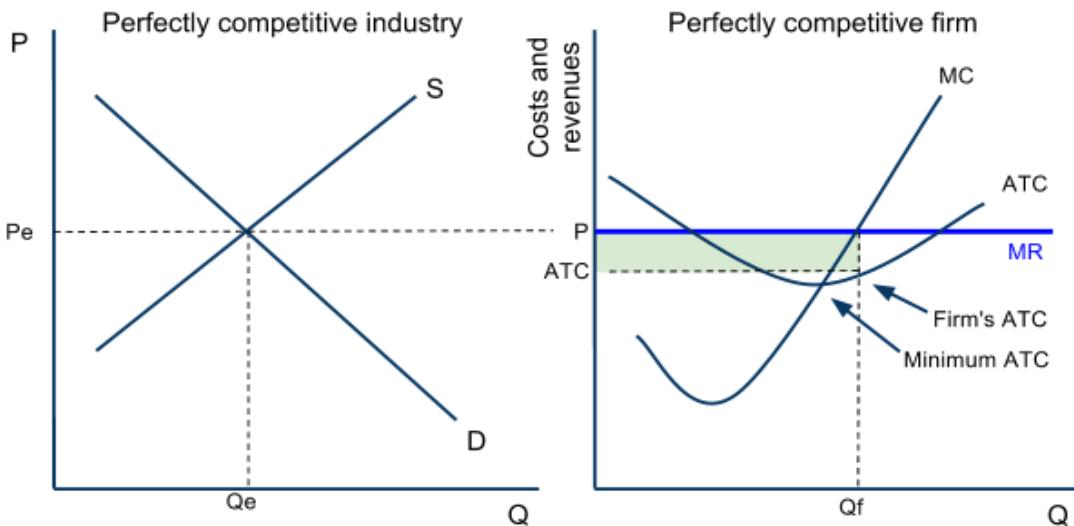
Productive efficiency

Price acts as a signal in competitive markets of the demands (and therefore the marginal benefit) of consumers. Price will always equal firms' minimum ATC in the long-run, assuring that perfectly competitive sellers will be productively efficient.

- If price is high enough that firms are earning profits, then the signal from buyers to sellers is WE WANT MORE
- If price is low enough that firms are earning losses, then the signal from buyers to sellers is WE WANT LESS

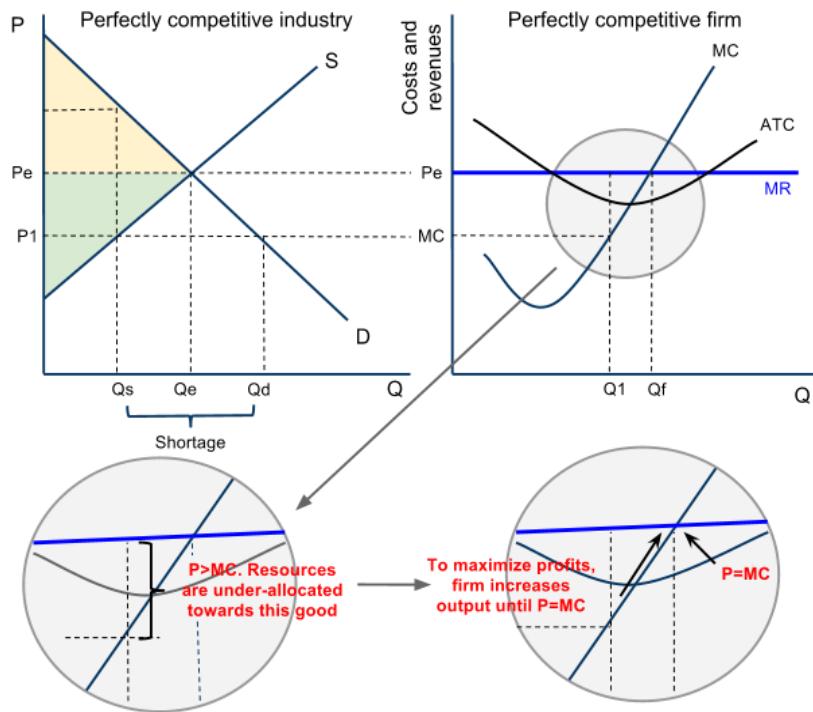
Consider the market and firm seen below:

- Price is higher than the firm's ATC.
- The firm's are earning economic profits
- The signal from buyers is “we want more”, so more firms will enter the market to satisfy demand.
- As new firms enter, price will fall to minimum ATC, and firms will be more **productively efficient!**



Study the graphs here:

- Assume the firm (which represents all firms in this market) produces at Q_1 .
- Market quantity supplied will be only Q_s . At Q_s , the demand (MB) is greater than the supply (MC) of the good. Resources are under-allocated.
- The profit-maximizing firm will increase its production to Q_f to achieve the $MR=MC$ point.
- As all firms do so, market quantity increases to Q_e .
- When all firms produce where $P=MC$, the shortage that existed at Q_s is eliminated and resources are efficiently allocated toward this good!



1.5.3 Monopoly

Introduction to imperfect competition

A market is **imperfectly competitive** if there are relatively few firms competing with one another either to sell products to consumers or to hire workers in a labor market. Common characteristics of imperfectly competitive markets include:

- Individual firms are **price-makers**: This means that they have the ability to raise or lower their product prices. In contrast, a perfectly competitive firm is a price-taker: it must sell its output for whatever the market price is.
- There are **barriers to entry and exit**: This means that in the existence of profits or losses firms will not immediately and effortlessly enter or exit the market. Firms may enjoy profits due to entry barriers or sustain losses for longer due to exit barriers.
- **Price will be higher than the firm's marginal cost**: Because entry barriers exist, firms are able to charge prices higher than their marginal costs. This also means that imperfectly competitive markets are allocatively inefficient: Market output will be at a level below the socially optimal level of output.
- Firms are **productively inefficient**: Because they are protected by entry barriers, firms will produce at an average cost that is higher than their minimum average total cost, meaning the firms are not using their resources to their maximum efficiency.

Not all imperfectly competitive markets are the same, of course. There is a range of imperfect competition from markets that are highly competitive with a large number of firms, to pure monopolies, in which a single firm produces all the output in the market.

- **Monopolistic competition**: A market with a large number of firms that sell similar (but not identical) differentiated products is monopolistically competitive. Firms have price-making power, but there are relatively low entry and exit barriers. In the long run, economic profits will attract new competitors and individual firms will only break even, unless they continue to differentiate and promote their products.
- **Oligopoly**: A market with a few dominant firms, each with a significant percentage of the total market share, is oligopolistic. There are high entry barriers protecting firms' economic profits, even in the long run. Because of the small number of firms, each firm tends to be interdependent on the actions of its competitors. For this reason, the price, output, and other decisions made by oligopolies must be made strategically, always considering what the competition will do in response to a particular decision.
- **Pure monopoly**: A market with a single firm producing nearly all or all of the market output is a monopoly. Such firms have total price making power: they can charge whatever price they choose as there are no substitutes to which consumers can switch. Entry to the market is virtually blocked, guaranteeing the monopolist's dominance. Some monopoly industries must be regulated by government, particularly those that produce essential goods or services such as energy utilities or transport networks.
- **Monopsony**: A firm that is a dominant or the sole employer of a particular resource,

such as labor, is a monopsony, which means “single buyer.” In labor markets, monopsonies have much control over the market wage rate, often paying a lower wage to its workers than would be paid in a more competitive labor market.

Assumptions of the monopoly model

- Describe, using examples, the assumed characteristics of a monopoly: a single or dominant firm in the market; no close substitutes; significant barriers to entry.

Monopolistic markets differ from perfectly competitive markets in nearly all characteristics. Study the table below to compare the two market structures.

Characteristic	Pure (or Perfect) Competition	Pure Monopoly
Number of Firms	VERY large number of firms	Only ONE firm. The firm IS the industry
Price making abilities of individual firms	Each firm is so small that changes in its own output do not affect market price, i.e. firms are price takers	Changes in the firm's output cause changes in the price, i.e. the firm is a price-maker!
Type of product	Firms all produce identical products, with no differentiation	Unique product, no other firm makes anything like it.
Entry barriers	Completely free entry and exit from the industry, i.e. NO barriers to entry.	Significant barriers to entry exist, preventing new firms from entering and competing with the monopolist
Efficiency	Will achieve both allocative and productive efficiency in the long-run	Will achieve neither allocative or productive efficiency in the long-run

Pure monopoly is a market structure in which there is only ONE dominant firm that sells a unique product, has price-making power and in which there are significant barriers to entry.

Monopoly in the real world: Monopolistic markets are, in fact, more common than perfectly competitive markets. Quite a few of the goods and services we consume are provided by pure monopolies or at least NEAR monopolies:

- Microsoft: has a near monopoly in the market for PC operating systems, in which its Windows software runs on nearly every PC computer in the world.
- Local utilities: Most of us have only one option from where we buy our electricity, water, garbage collection, and gas. Most public utilities are provided by monopolists
- State liquor stores: In many US states liquor is sold in purely monopolistic state-run (or regulated) stores
- Cable and phone providers: Until the last decade or two, most people had only one option from where to buy their cable TV or their phone service. The adoption of cellular phone technology has made the phone service industry more competitive recently.

- Rail transportation: In the US, Switzerland, and many other countries, there is a purely monopolistic provider of train service in the country. If you want to travel by train across the US, you will travel on Amtrak.

Barriers to entry

- Explain, using examples, barriers to entry, including economies of scale, branding and legal barriers.

One characteristic all monopolies share is that there are significant barriers to entry, which keep competition out of the market. It is these entry barriers that protect a monopolist's power. Without high entry barriers, new firms would enter the market and reduce the price-making and profit-making power of the monopolist.

Examples of entry barriers:

- Legal barriers: Monopolists may have exclusive rights granted by the government to provide a certain good or service. Other legal barriers may include patents or copyrights held by the firm that prevent competition from producing a similar product.
- Economies of Scale: The “advantages of being big”. Some firms have achieved such a great size that they can simply produce their good more efficiently, and thus sell it for a lower price, than any other firm could hope to do, keeping competition out of the market.
- Ownership of resources: If a firm has exclusive access to the resources needed to make its good, then no other competitor can hope to begin producing the good. An example of this is the global diamond giant De Beers, which has exclusive access to over 80% of the known diamond mines in the world.
- Strategic pricing: A monopolist may be able to block entry to the market by temporarily selling its output at a price below its per-unit costs (and earning short-run losses). This deters competitors from entering
- Brand loyalty: If a firm has a brand that is well known and popular among consumers, then other firms will find it hard to get a foothold in the market, allowing the monopolist to maintain market share.

Revenue curves for the monopolist

- Explain that the average revenue curve for a monopolist is the market demand curve, which will be downward sloping.
- Explain, using a diagram, the relationship between demand, average revenue and marginal revenue in a monopoly.
- Explain why a monopolist will never choose to operate on the inelastic portion of its average revenue curve.

Demand, average revenue and marginal revenue as seen by the monopolist are quite different as that seen by the perfectly competitive firm.

- Recall that in perfect competition, demand, MR, and AR as seen by the firm is a horizontal line equal to the equilibrium price determined in the market.

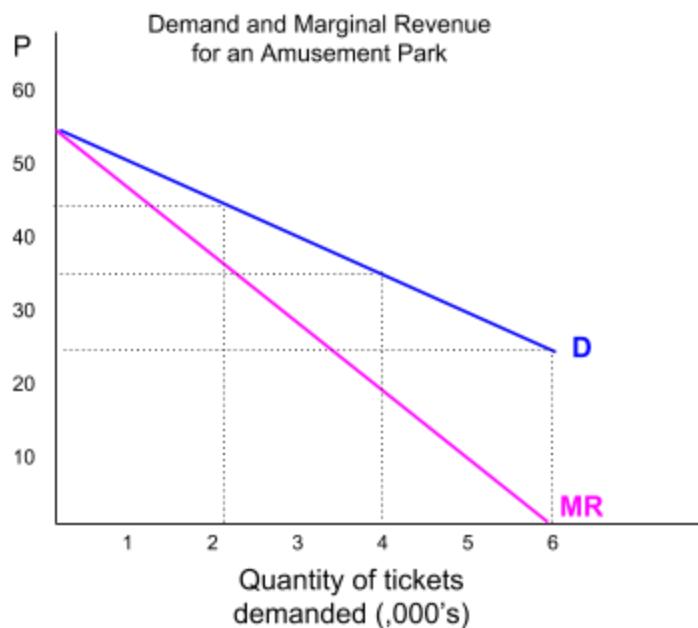
- In monopoly, the demand seen by the firm is the market demand, and MR falls faster than demand, AR and price.

The table below represents the demand schedule, the total revenue, and the marginal revenue of an amusement park (the only amusement park in a region: in other words, a monopolist).

Q (thousands)	P (AR)	TR (PxQ) thousands	MR ($\Delta TR/\Delta Q$) thousands
0	55	0	-
1	50	50	50
2	45	90	40
3	40	120	30
4	35	140	20
5	30	150	10
6	25	150	0

Observe the following about the monopolist's demand and revenues:

- At \$55, no tickets will be sold. At \$50, 1,000 will be sold. In order to sell more tickets, the park must lower prices. The park is a price-maker!
- The park's revenues rise until it has sold 5,000 tickets, where it peaks at \$150,000.
- MR falls as output increases, but it falls twice as rapidly as the price.
- Graphically, the MR will be below the demand curve.



Observe from the graph:

Because the monopolist must lower its price to sell additional units, its marginal revenue of a particular unit will always be lower than the price that unit sells for (except at an output of 1).

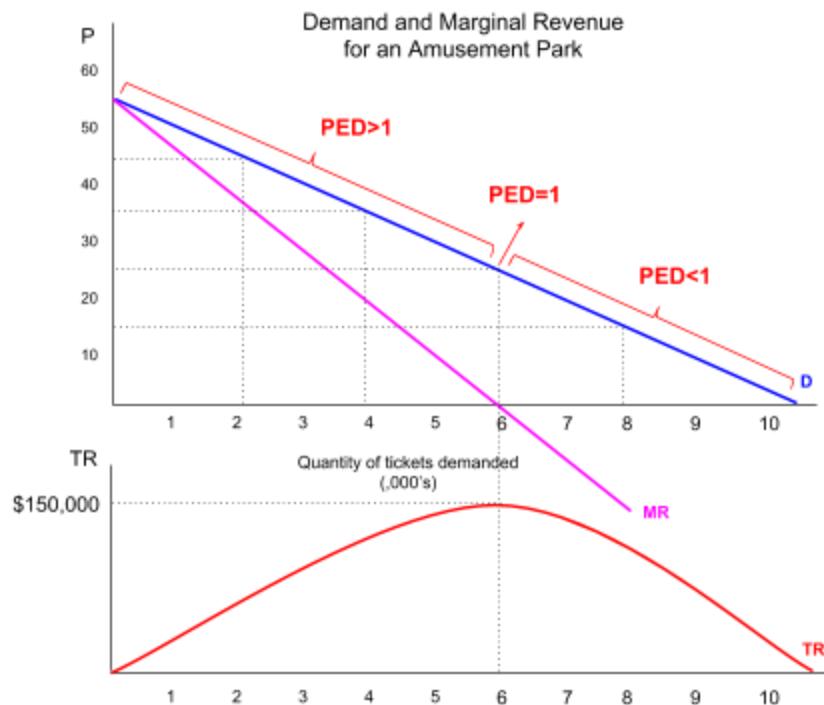
Points about the monopolist's demand:

- Demand for the firm's output IS the market demand
- Demand is relatively inelastic compared to a perfectly competitive firm, since there are no close substitutes for a monopolist's product.
- When MR is positive, lowering the price and increasing output will cause the firm's total revenues to rise.
- When MR is negative (beyond 6 units on this graph), further price decreases and quantity increases will cause the firm's total revenues to fall.

PED and the monopolist's demand

In a competitive market in which hundreds of firms producing identical products compete with one another, demand for each firm's output is perfectly elastic, meaning no firm is able to raise its price as buyers will switch to another seller. A monopolist, on the other hand, is a price-maker. The firm has the freedom to raise or lower its price (and change its level of output) at will, since it is the sole producer of the good.

Compared to a perfectly competitive seller, demand for a monopolist's product is relatively inelastic. However, as a price-making firm lowers its price, the PED for its good decreases, while at higher prices PED increases. To understand why, consider the graph below.



Observe from the graph:

- As this firm increase its output from 0 to 6,000 units, it must lower its price from \$55 to \$25 in order to sell more output.
- As it does so, its total revenues ($TR=P\times Q$) increase from \$0 to \$150,000.
- At a quantity of 6,000 and a price of \$25, the firm's TR is maximized.
- In order to sell more output, the firm must lower the price, but since consumers are relatively unresponsive to further price cuts the firm's revenues fall beyond 6,000 units.

You will recall from the unit on elasticities that if a decrease in price causes sellers' revenues to increase, demand is elastic, and when a decrease in price causes sellers' revenues to decrease, demand is inelastic. The monopolist's (or any price-making firm), demand can therefore be observed to have:

- An elastic range where MR is positive. Consumers are relatively responsive to lower prices, so the percentage increase in Q_d is greater than the percentage decrease in price.
- An inelastic range where MR is negative. Consumers are relatively unresponsive to lower prices, so the percentage increase in Q_d is lower than the percentage decrease in price.
- A price and quantity combination at which demand is unit elastic where $MR=0$. If a change in price causes no change in total revenue (in other words, TR is maximized), then the percentage increase in Q_d is equal to the percentage decrease in price.

A price-making firm will NEVER produce in the inelastic range of its demand! Because if a monopolist were to sell beyond the point where $MR=0$, it would always do better by decreasing its output until MR were positive once again.

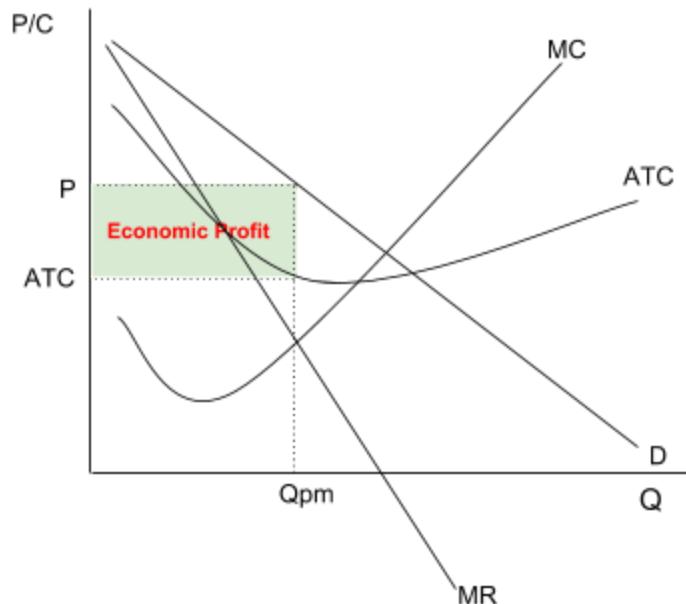
- Total costs would decrease as the firm reduces its output
- Total revenue would increase, therefore...
- Reducing output to a point below the revenue maximizing quantity (where $MR=0$) would definitely increase the firm's profits (remember, economic profits = $TR-TC$)

Profit maximization in monopoly

- Explain, using a diagram, the short- and long-run equilibrium output and pricing decision of a profit maximizing (loss minimizing) monopolist, identifying the firm's economic profit (abnormal profit), or losses.
- Examine the role of barriers to entry in permitting the firm to earn economic profit (abnormal profit).

Just like a firm in perfect competition, a monopolist wishing to maximize its profits wants to produce at the quantity at which Marginal Revenue (MR) = Marginal Cost (MC).

To determine a monopolist's profit maximizing level of output, therefore, we must consider both its revenues and its costs. The graph below shows a monopolistic firm producing at its profit maximizing price and quantity, where $MR=MC$.



Notice in the graph:

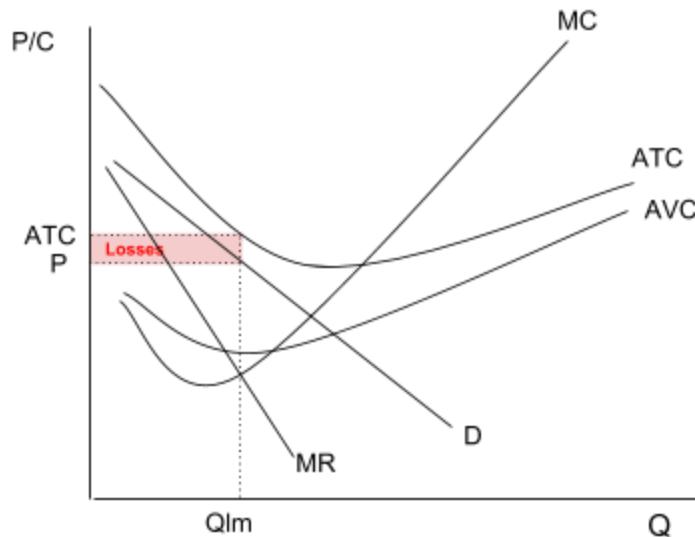
- The monopolist's MC and ATC demonstrate the same relationships as a firm in perfect competition.
- The firm will produce at the quantity at which $MC=MR$ to maximize profits.
- Subtracting the firm's ATC at Q_{pm} from the price it can sell Q_{pm} units of output for, and multiplying by the quantity produced find the area of economic profit.
 $Economic\ Profits = (P-ATC) \times Q$.
- Because of the entry barriers in this market, the firm's profits are sustainable in the long-run

The loss minimizing monopolist

Having monopoly power does not guarantee that a firm will earn economic profits.

- If demand for a monopolist's output falls, or
- If the monopolist's costs of production rise, then...
- The firm can go from earning economic profits to earning losses.

To minimize losses, a monopolist should produce at its $MR=MC$ level of output.

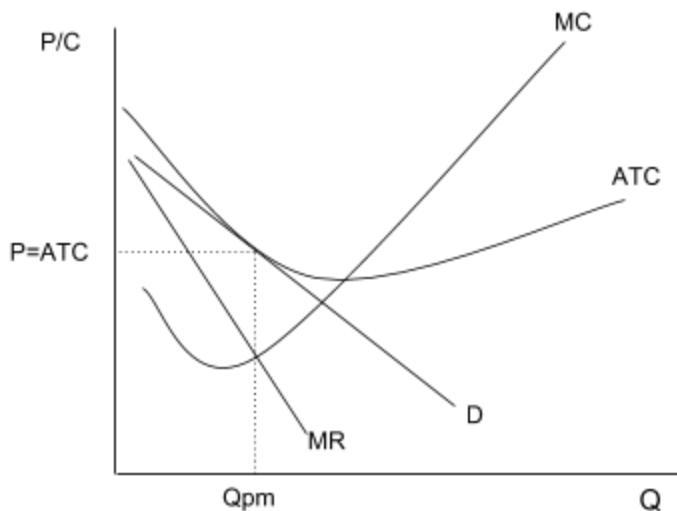


Notice in the graph:

- The firm is producing at its $MR=MC$ level of output, but at this point the firm's ATC is greater than the price it can sell for.
- The firm is earning economic losses represented by the triangle.
- Despite its losses, this firm should NOT SHUT DOWN, because the price still covers the average variable cost; this firm can continue to operate in the short-run.
- Only if total losses were larger than the total fixed costs should the firm shut down.
- To reduce or eliminate its losses, the firm must try and increase demand or reduce its costs.

The breaking-even monopolist

Of course, it is also conceivable that a monopolist will be selling its product at a price that is exactly equal to its ATC . This would mean that the monopolist is breaking even.



Notice in the graph:

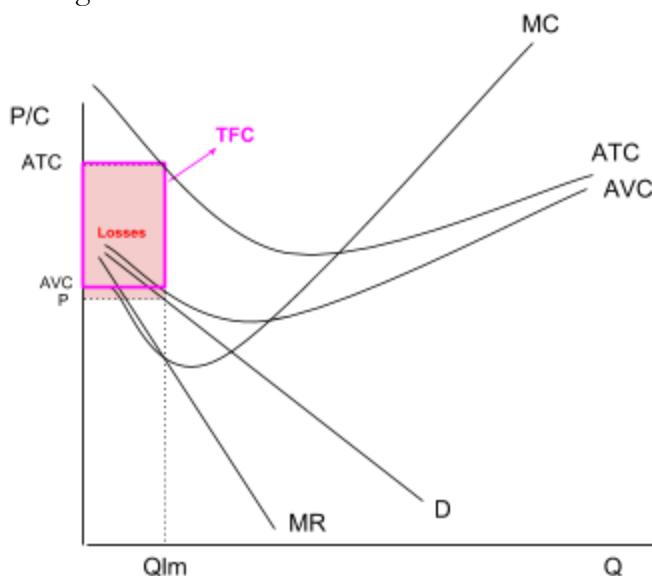
- The firm is producing at its $MR=MC$ level of output. At this point the firm's ATC is exactly equal to its price.

- The firm's total revenues are exactly equal to its total costs.
- The firm is covering all of its explicit and implicit costs, meaning it's earning a normal profit, but it is not earning any economic profit.
- If the firm wishes to earn economic profits, it will have to improve or advertise its product to increase demand or increase the efficiency with which it uses resources to reduce its costs.

When a monopolist should shut down

Firms should follow a simple rule when deciding whether or not to shut down and leave a market:

- If the price it can sell for is lower than the firm's average variable cost, or...
- If total losses are greater than the firm's total fixed costs.



Notice in the graph:

- The firm is producing at its $MR=MC$ level of output, but at this point the price the firm can sell its output for is lower than the firm's average variable cost.
- This firm cannot even afford to pay its workers for each unit they produce (the per-unit labor costs are higher than the price)
- The gray rectangle represents the firm's losses $(ATC-P) \times Q$. The firm's total fixed costs, $(ATC-AVC) \times Q$, are smaller than the total losses. This means that if the firm shuts down it will minimize its losses

Long-run equilibrium in a monopolistic market

In our study of perfect competition we learned the following:

In perfectly competitive markets

- If firms are earning economic profits in the short-run, new firms will enter the market, increasing the supply, reducing the price and eliminating profits.
- If firms are earning economic losses in the short-run, some firms will exit the market, reducing the supply, increasing the price and eliminating losses for the firms

that remain.

- In the long-run, firms in perfectly competitive markets will only break even.

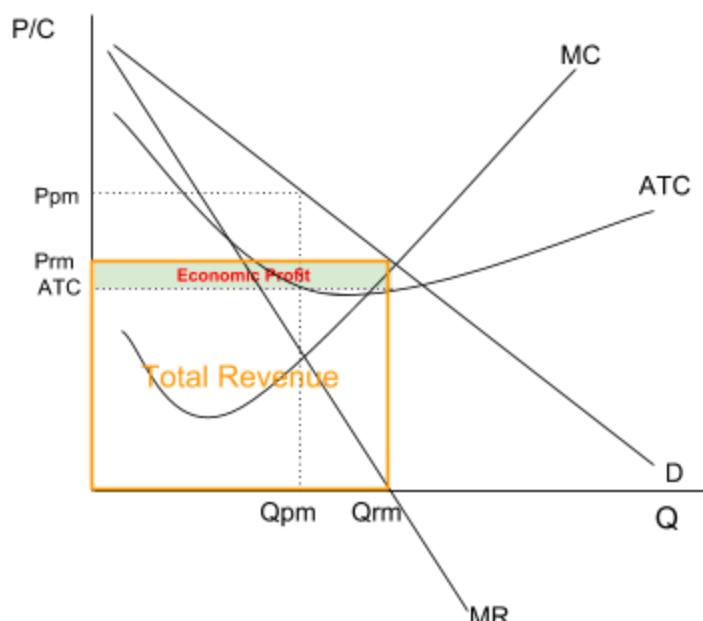
However, in monopolistic markets:

- If the firm is earning economic profits in the short-run, those profits will be maintained as long as the firm can keep demand for its goods high and its costs low, because entry to a monopolistic market is blocked!
- If the firm is earning economic losses in the short-run, those losses will be maintained as long as the firm cannot increase the demand for its product or reduce its price. Exit from a monopoly market is difficult because of the large economies of scale that often characterize large, single sellers.

Revenue maximization

- Explain, using a diagram, the output and pricing decision of a revenue maximizing monopoly firm.
- Compare and contrast, using a diagram, the equilibrium positions of a profit maximizing monopoly firm and a revenue maximizing monopoly firm.
- Calculate from a set of data and/or diagrams the revenue maximizing level of output.

A price-making firm may not always aim to maximize profits, rather to sacrifice profits for market share and higher revenue. The firm below is earning smaller profits at Q_{rm} and P_{rm} (**revenue maximizing** quantity and price) than it would at Q_{pm} and P_{pm} (where $MC=MR$ and profits are maximized), however, it sells more output and may therefore have more market share than it would at the higher price.



The revenue maximizing firm:

- Produces where the last unit sold earned the firm no additional revenue (where $MR=0$)

- At every level of output up to this point increases in output caused revenues to rise ($MR>0$)
- At every level of output beyond this point increases in output caused revenues to decrease ($MR<0$)
- Total revenues are maximized where $MR=0$, but the level of economic profit is smaller than it would be at a lower quantity and higher price.

Natural monopoly

- With reference to economies of scale, and using examples, explain the meaning of the term “natural monopoly”.
- Draw a diagram illustrating a natural monopoly.

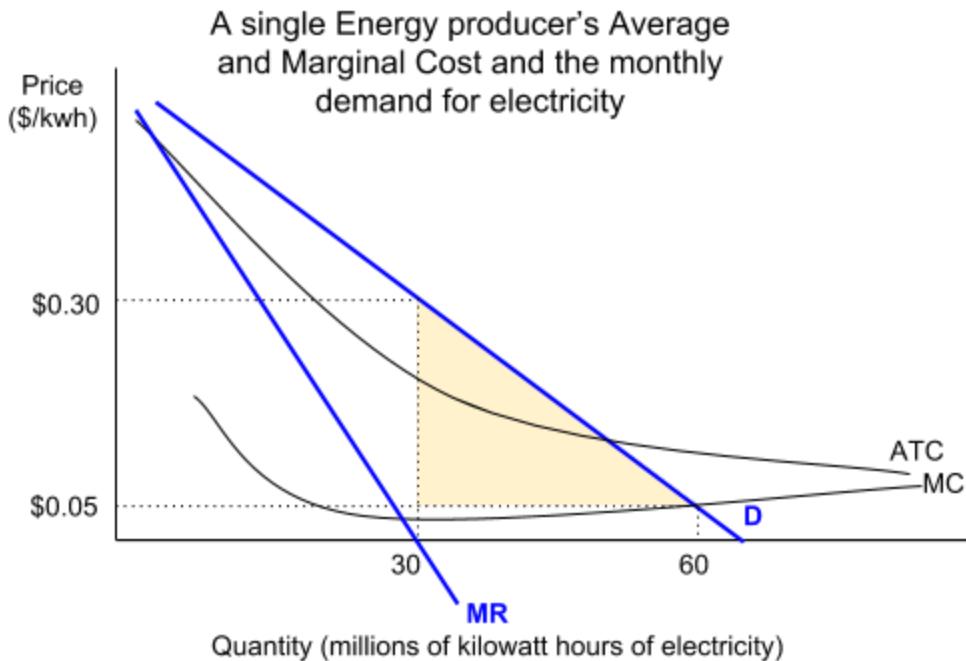
Not all industries that are monopolies necessarily need to be monopolies. In other words, sometimes firms have monopoly power for legal or technical reasons:

- If a firm has an exclusive permit from the government to provide a particular good
- If a firm has “cornered the market” for a particular resource needed to produce the good
- If a firm has priced competitors out of the market using predatory pricing strategies...
- Any of these sources of monopoly power could be considered economically inefficient and therefore undesirable to some extent.

However, there is a type of monopolistic industry in which the dominance of a single firm is economically justifiable and actually beneficial for society! It is called a **natural monopoly**.

A monopoly in a key industry like electricity generation can potentially be very harmful for consumers of the product being produced. Monopolists tend to charge a price that is higher, and produce a quantity that is lower than what is socially optimal.

A **natural monopoly** is an industry in which a single firm produces all the output because of its large economies of scale. Demand for the good is low relative to the very high fixed costs. In the graph below we see a natural monopoly, recognized because demand intersects ATC while it is still in its “increasing returns to scale”, downward sloping range.



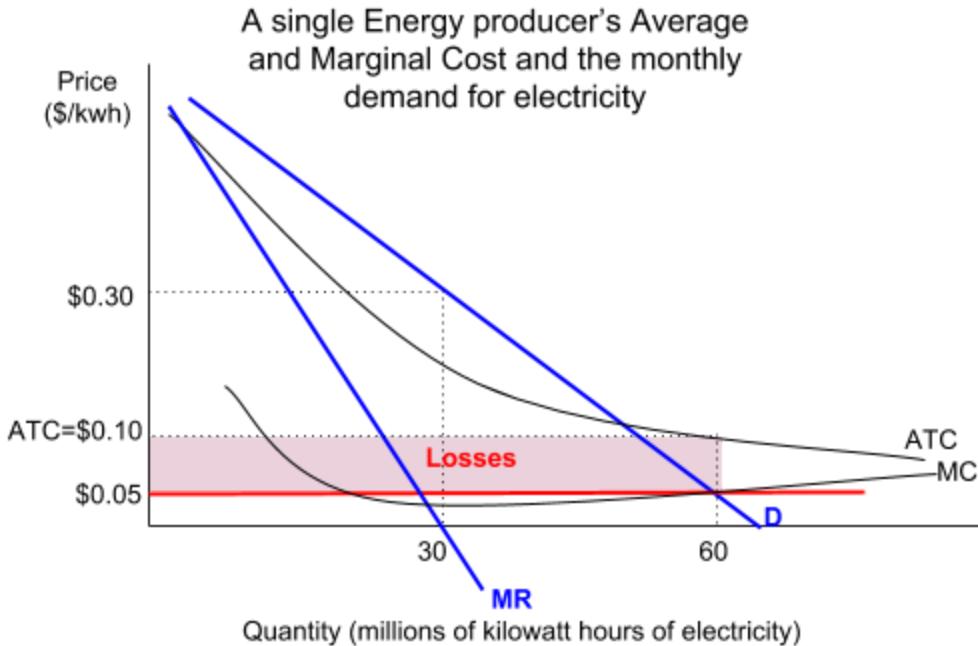
The diagram above shows the monthly demand for electricity in a small city and the average and marginal costs of the naturally monopolistic utility company that provides the city with its electricity.

- If the firm is left to produce at its profit-maximizing price and quantity, it will limit its production to 30 million kwh and charge a price of \$0.30 per kwh.
- The socially optimal quantity of electricity (where $P=MC$ or marginal benefit equals marginal cost) is 60 million kwh and the socially optimal price is \$0.05 per kwh.

Clearly, the utility company, if left to pursue its profit-maximizing agenda, will overcharge for and under produce electricity. There is a loss of consumer surplus represented by the yellow triangle as a result of the monopoly abusing its power.

To ensure a more socially optimal level of output and price, government regulation is needed. Either subsidies or price ceilings (or both) will increase output and reduce price.

To ensure that a naturally monopolistic industry produces at a level closer to the allocatively efficient or socially optimal price and quantity (where $P=MC$), either subsidies or price controls should be imposed by the government.



The government could establish a price ceiling at \$0.05 per kwh. As we learned previously, when a price ceiling is set at the socially optimal price, a monopolist will increase its output to the socially optimal level (60 million kwh in this case).

There is one problem with a price ceiling set at the socially optimal price in a natural monopoly, however. It will always be at a price lower than the firm's ATC, meaning the firm will experience economic losses. The red area on the graph above represents this utility company's losses when it is forced to charge the socially optimal price of \$0.05. At this price, the company will choose to shut down in the long-run.

The solution, therefore, is for the government to provide the firm with a **lump-sum subsidy** to cover the losses it experiences when producing at the socially optimal level. In the case of our utility company, the subsidy would have to equal at least the level of losses the firm experiences at \$0.05 per kwh.

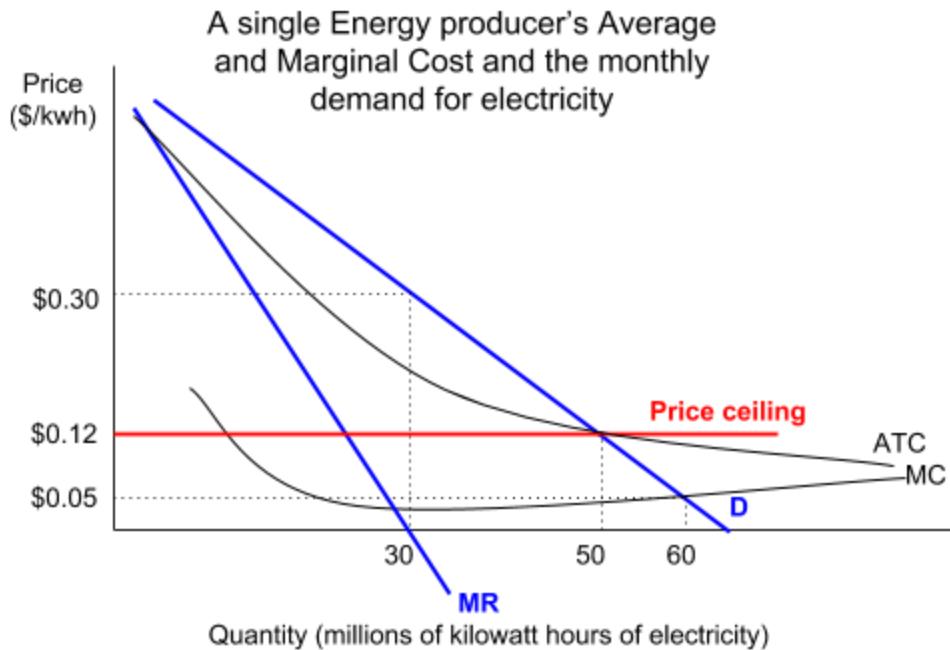
$$\begin{aligned}
 \text{Economic losses} &= (\text{ATC}-\text{P}) \times \text{Q} \\
 &= (\$0.10 - \$0.05) \times 60 \text{ million} \\
 &= \$0.05 \times 60 \text{ million} \\
 &= \mathbf{\$3 \text{ million}}
 \end{aligned}$$

A \$3 million lump-sum subsidy (payment) from the government to the utility company will assure it is able to stay in operation even while charging the socially optimal price of \$0.05 per kwh and producing the socially optimal quantity of 60 million kwh.

The “fair return” price ceiling

Another option for regulating natural monopolies that would not involve providing

lump-sum subsidies is to establish what is called a “**fair return**” **price ceiling**. This is a price ceiling set at the price where demand crosses the firm’s ATC, as seen in the graph below.



Observe the following in the graph above:

- A price ceiling of \$0.12, which is equal to the firm’s average total cost, will increase quantity demanded to 50 million kwh and make electricity more affordable, promoting a more socially optimal level of production and price
- Since the firm breaks even, no government subsidy is necessary
- This is called a “fair return” price because the firm charges a price that earns it a normal profit. No economic profit is earned at the expense of consumers who would be unable to afford electricity.

Subsidies and price controls are two commonly used interventions in naturally monopolistic utility industries such as electricity, water, gas, public transportation, waste management and recycling, etc... Capping the price the monopolistic producers are allowed to charge at or closer to the socially optimal price (where MB-MC) and using subsidies to cover any resulting losses will assure that essential services are provided at a relatively low cost and at a relatively high level to society.

Monopoly and efficiency

- Explain, using diagrams, why the profit maximizing choices of a monopoly firm lead to allocative inefficiency (welfare loss) and productive inefficiency.
- Evaluate reasons why, despite inefficiencies, a monopoly may be considered desirable for a variety of reasons, including the ability to finance research and development (R&D) from economic profits, the need to innovate to maintain economic profit (abnormal profit), and the possibility of economies of scale.

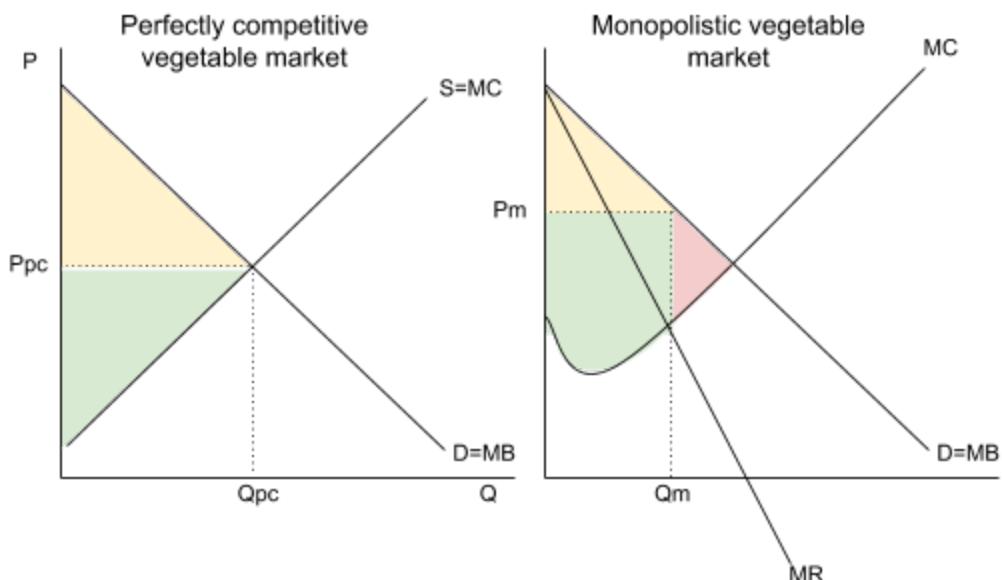
- Draw diagrams and use them to compare and contrast a monopoly market with a perfectly competitive market, with reference to factors including efficiency, price and output, research and development (R&D) and economies of scale.

As we learned in the previous unit, perfectly competitive industries are both allocatively and productively efficient. This is because, in the long-run:

- Price will always equal marginal cost (the allocatively efficient level of output)
- Firms will always produce at their minimum ATC (the productively efficient level of output)

To determine whether monopolies are efficient, we must consider whether the same conditions are met.

Allocative efficiency: Consider the two vegetable markets below. In one, hundreds of small vegetable growers compete to provide the market with fresh produce. In the other, a monopolistic seller controls the market price and quantity to maximize its own profits.

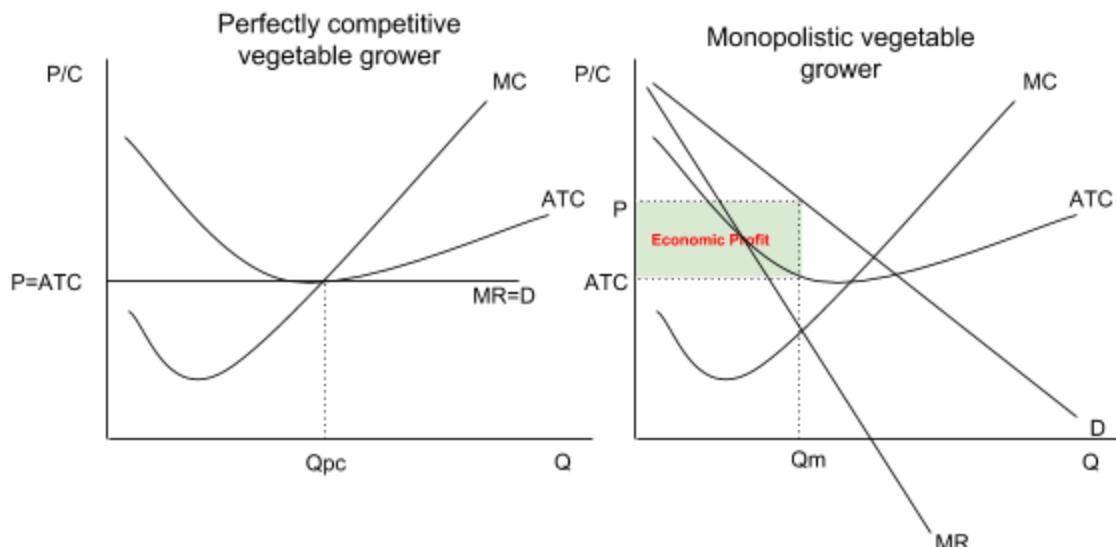


Observe from the graphs:

- In the perfectly competitive market, output is determined by market supply and market demand, therefore the socially optimal price and quantity are established (where $MB=MC$)
- Since the price of vegetables equals the marginal cost, the market is allocatively efficient.
- In the monopoly market the single producer limits output to Q_m , where $MR=MC$, in order to maximize its profits.
- Consumers pay a higher price and have less vegetables to choose from.
- Since $P>MC$, resources are underallocated towards vegetables, and there is a welfare loss (DWL) represented by the red triangle.

Productive efficiency: Under perfect competition, firms are forced to be productively efficient, meaning they produce their products at the lowest possible average total cost. Without competition, however, monopolists are NOT productively efficient.

Consider the two firms:



Observe from the graph:

- The vegetable grower producing in the competitive market is forced to produce at its minimum average total cost, at which the grower breaks even. At any other level of output the perfect competitor would earn losses. Competition forces producers to use resources in their most efficient way, maximizing productive efficiency.
- The monopolist, which faces no competition, produces at a quantity lower than where its ATC is minimized and charges a price higher than its minimum ATC. The lack of competition allows the monopolist to use resources less efficiently and focus on maximizing its profits by restricting output and charging a higher price.

Compared to perfect competition, monopolistic markets have several observable effects

Compared to competitive markets monopolies will demonstrate...
Higher price
Lower output
$P > \text{min. ATC}$: Productive inefficiency
$P > MC$: Allocative inefficiency (resources are under-allocated towards the product)
Efficiency Loss (Welfare loss) occurs

A loss of Consumer surplus in exchange for higher firm profit. Welfare loss results
Income transfer: consumers pay a higher price, shareholders of the monopoly enjoy higher profits.

Some other effects of an industry becoming a monopoly include:

- Economies of scale: Some monopolized industries have only one firm because economies of scale exist over such a wide range of output. It is possible that one or two large firms can achieve a lower ATC than many smaller firms. This is called a natural monopoly.
- Simultaneous consumption: One product can satisfy a large number of consumers at the same time. Example: Microsoft Windows. Marginal Cost for Microsoft is essentially nothing, so ATC_{LR} declines over the entire range of output.
- Network effect: describes the phenomenon of a product's value increasing the more users it has. Examples: cell phones, the internet, email, Facebook! Tends to move markets towards monopoly as more and more consumers flock to a product because of the "network" that develops around it.
- Income Transfer: Consumer surplus is lost because of higher price. Firm profits are higher b/c of market power. Compared to PC industries, monopolies represent a transfer of income from consumers to shareholders in the monopolistic firm.

1.5.4 Monopolistic Competition and Oligopoly

Assumptions of the monopolistic competition model

- Describe, using examples, the assumed characteristics of a monopolistic competition: a large number of firms; differentiated products; absence of barriers to entry and exit.

Introduction to monopolistic competition

The third market structure we will study gets its name from sharing some characteristics with pure monopoly and some with perfect competition. Below are some of the key characteristics of the monopolistically competitive market:

Characteristic	Monopolistic Competition
Number of Firms	Fairly large number of firms, each with a relatively small amount of market share
Price making abilities of individual firms	Firms are small relative to the industry, meaning changes in one firm's output have only a slight impact on market price. While they are price-makers, demand will be relatively elastic compared to a pure monopolist
Type of product	Products are slightly differentiated. Firms will advertise to try and further differentiate product. Branding and advertising are used to attempt to increase demand for the firm's product over competitors.
Entry barriers	Entry to and exit from the market is relatively easy. If profits exist, new firms will enter, if losses are earned, it can be expected that some firms will exit.
Efficiency	Because of their price-making power, firms will produce at a price that is higher than their marginal cost and higher than their minimum ATC, meaning the industry is not economically efficient.

Examples of monopolistically competitive markets

Monopolistic competition is probably the most common market structure in most market economies. The characteristics apply to a wide range of industries in which many sellers compete for the business of buyers. Examples include:

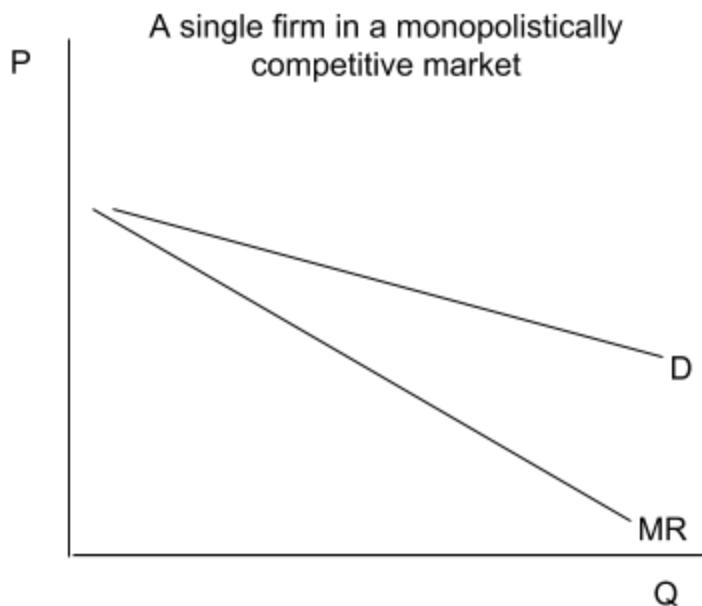
- Restaurants in a major city: There are hundreds of restaurants in a city of any reasonable size. They all sell a similar product (food), which is differentiated from one seller to the other (Chinese, Mexican, French, Barbecue, etc...) Each restaurant can set its own prices, but only to an extent (have you ever seen a \$100 hamburger?)
- Apparel: The market for clothing is highly competitive, and like restaurants, the hundreds (or thousands) of clothing manufacturers are competing for our business by differentiating their products from the competition. Again, firms have some price-making power, but consumers can always switch brands if prices rise too much, so demand is relatively elastic.
- Automobiles: Even the car market shows some characteristics of monopolistic

competition, although due to the relatively substantial economies of scale, it could be considered oligopolistic in some markets. Each car is a close substitute for all other cars, but is differentiated to try to make demand for it less elastic.

Revenue curves for the monopolistic competitor

- Explain that product differentiation leads to a small degree of monopoly power and therefore to a negatively sloping demand curve for the product.

Because each firm in a monopolistically competitive market makes a product that is differentiated from its competitors, it is able to control the price for its output, but only to a certain extent.



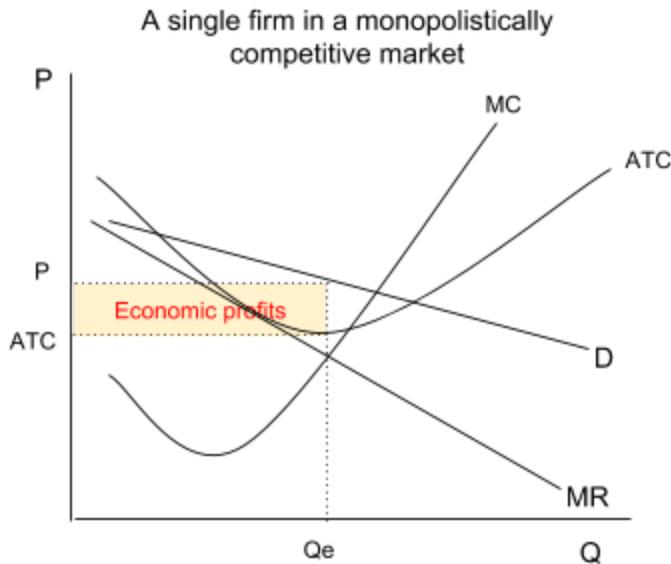
Observations of the monopolistic competitor's demand and MR curves:

- With many other firms making similar products, each firm faces a relatively, but not perfectly, elastic demand curve. A price increase will lead to a large loss of buyers, but a price decrease will lead to a large increase in buyers.
- In order to sell additional units of its product, a firm must lower the price of all its output. For this reason, the firm's marginal revenue will fall faster than its price (see a mathematical explanation for this in the chapter on Monopoly).

Profit maximization in the short run

- Explain, using a diagram, the short-run equilibrium output and pricing decisions of a profit maximizing (loss minimizing) firm in monopolistic competition, identifying the firm's economic profit (or loss).

As with firms competing in the other market structures, a monopolistic competitor will maximize its total profits when it produces at the quantity of output at which $MR=MC$



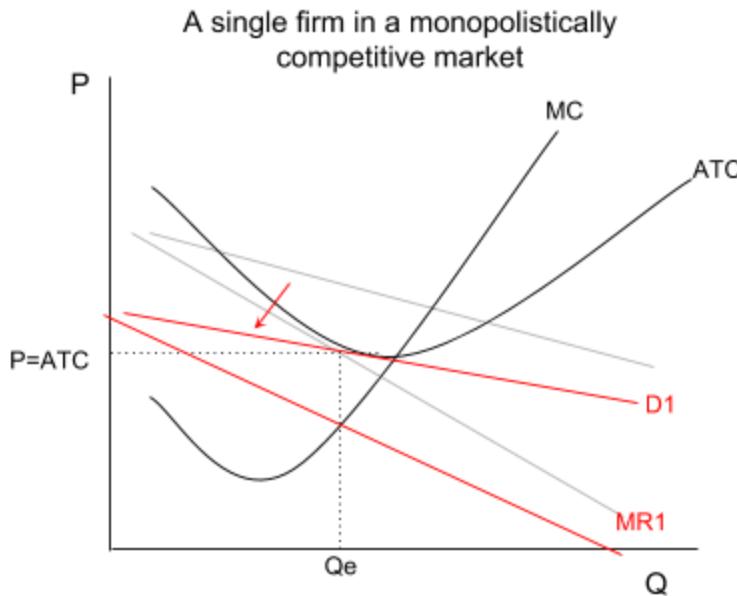
Observe from the graph:

- The firm is producing at its profit maximizing quantity (Q_f) and charging the price consumers are willing to pay for that quantity (P_f)
- At this point, price is greater than ATC, so the firm is earning an economic profit.
- Given the existence of profits in this market (assuming this firm is a typical firm) new firms will be attracted to the industry.
- Since entry barriers are low, these short-run economic profits are likely to be eliminated in the long-run as new firms enter the market.

Profit maximization in the long run

- Explain, using diagrams, why in the long run a firm in monopolistic competition will make normal profit.

One of the key characteristics of monopolistic competition is the low entry barriers. Getting into such a market is relatively cheap and easy, and entrepreneurs will therefore be attracted to any economic profits that are earned.

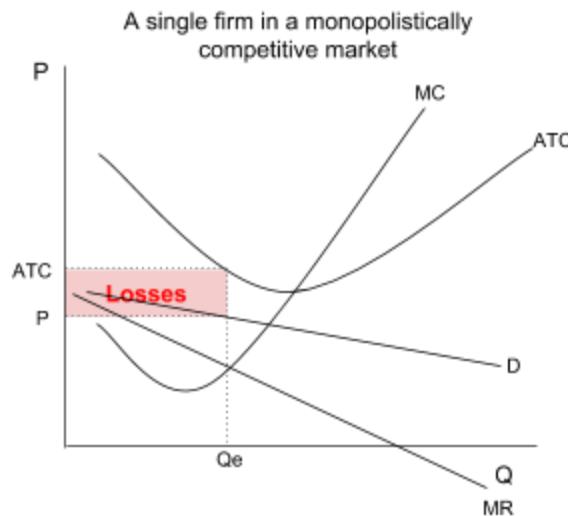


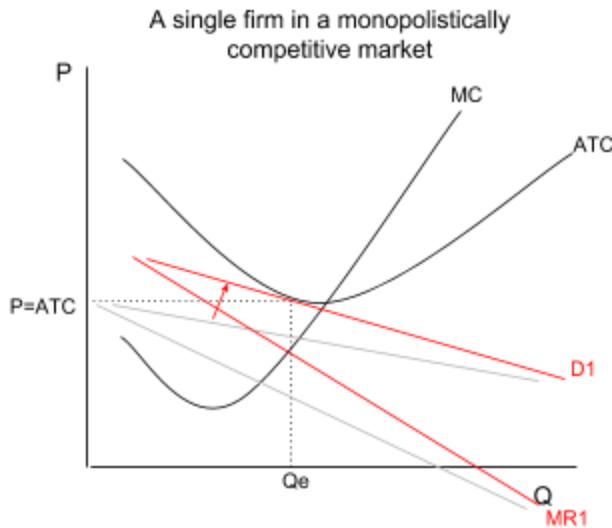
From short-run to long run in monopolistic competition:

- Economic profits attract new firms to the market, increasing the amount of competition and the number of substitutes for this firm's product
- More competition reduces demand for this firm's product, and makes it more elastic (flatter). Demand decreases until the firm is only breaking even

Notice that the demand curve is tangent to the ATC curve when the market is in long-run equilibrium. This means that at the $MC=MR$ point, the firm is only breaking even. The monopolistically competitive firm cannot do better in the long-run than break even, since any profits will be eliminated as new firms enter the market.

Just as it is relatively easy to enter a monopolistically competitive market, it is also easy to leave. This means that if the firms in such a market are earning losses, some will exit the market, increasing the demand for those that remain until they are breaking even.





In the graphs above:

- Due to weak demand or an increase in costs, firms are earning losses.
- Losses will lead some firms to exit the market, increasing market share for the remaining firms.
- In the second graph, demand has increased for a firm that has remained in the market. Due to fewer substitutes, demand also becomes more inelastic.
- Exit eliminates the losses and the remaining firms are once again breaking even.

Because of the low entry and exit barriers, firms in monopolistically competitive markets will only break even in the long-run (just like in perfect competition).

Non-price competition

- Distinguish between price competition and non-price competition.
- Describe examples of nonprice competition, including advertising, packaging, product development and quality of service.

Because firms face so much competition for their output, they will find it difficult to compete on price. In order to break even (or earn profits), a firm must compete through other, non-price means, including:

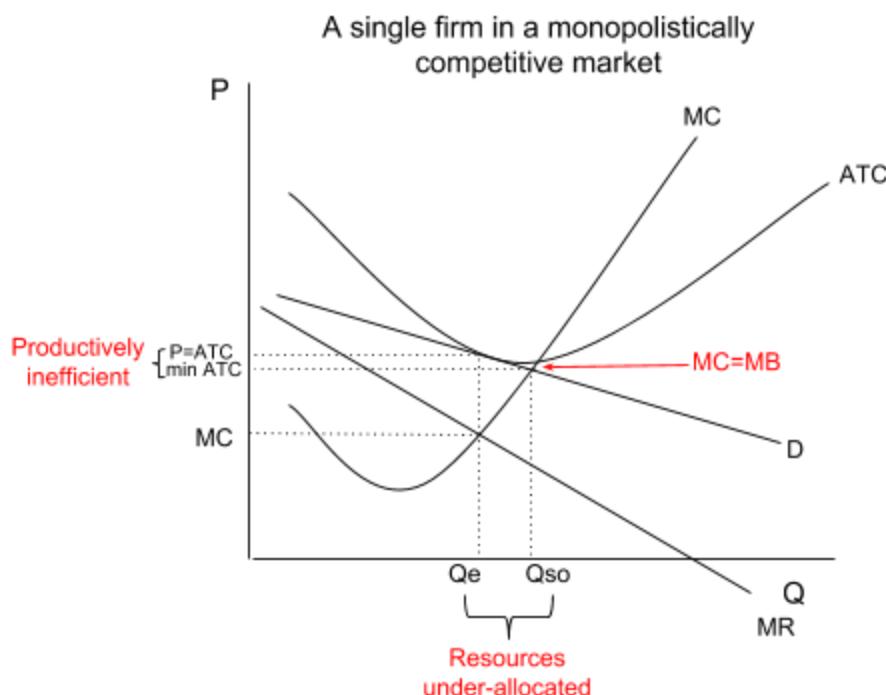
- Branding: By developing a recognizable brand image, firms attempt to build consumer loyalty, giving the firm more price-making power
- Product development: Continuously improving its product through research and development will keep demand high.
- Customer service: Offering good customer service and support may increase demand
- Location: Good access to large numbers of consumers allows a firm to charge higher prices
- Advertising: Making buyers aware of product features through advertising increases demand, giving the firm a greater chance of earning economic profits in the long-run

Monopolistic competition and efficiency

- Explain, using a diagram, why neither allocative efficiency nor productive efficiency are achieved by monopolistically competitive firms.

To determine whether monopolistically competitive firms are economically efficient, we must determine whether:

- $P = MC$: This is an indicator of allocative efficiency, since price represents the marginal benefits of consumers and MC the marginal cost to producers
- $P = \text{minimum ATC}$: This tells us whether firms are productively efficient, since if the price equals the lowest ATC, then firms are forced to use their resources in the least-cost manner.



Efficiency is not achieved! As we can see in the graph, a monopolistic competitor in long-run equilibrium will achieve neither productive nor allocative efficiency. The lack of competition allows firms to produce at a cost higher than their minimum ATC and produce a quantity lower than what is socially optimal.

Monopolistic competition compared with perfect competition and monopoly

- Compare and contrast, using diagrams, monopolistic competition with perfect competition, and monopolistic competition with monopoly, with reference to factors including short run, long run, market power, allocative and productive efficiency, number of producers, economies of scale, ease of entry and exit, size of firms and product differentiation.

It may appear that, since they do not achieve economic efficiency, monopolistically

competitive markets are less desirable than perfectly competitive markets. However, there are also several benefits of monopolistic competition over perfect competition.

Characteristic	Perfect Competition	Monopolistic Competition
Price and Quantity	Price is low and quantity is high. Allocative and productive efficiency are achieved and consumer surplus is maximized as a result.	Price is higher and quantity lower than in perfect competition, neither type of efficiency is achieved and consumer surplus will be less.
Product Variety	Every firm sells an identical product. There is no variety for consumers to choose from.	Every firm differentiates its product, at least slightly, from every other seller, giving consumers a wide variety to choose from.
Profits	Firms will always break even in the long-run, and due to the high level of competition there is nothing an individual firm can do to earn profits, only an increase in market demand can lead to short-run profits	Firms have more ability to make profits through successful non-price competition and product differentiation, which if done well can earn a firm profits, even over time.

If we compare monopolistically competitive markets to monopolistic markets, we observe some key differences as well.

- Lower entry barriers mean monopolistically competitive markets will have more firms competing for consumers
- More competition means greater efficiency and lower prices
- More competition means lower economic profits (which may mean less profits to invest in innovation product development) in monopolistic competition compared to monopoly.
- Differentiated products mean consumers have more choice than under pure monopoly, in which a single firm produces a unique product.
- A single monopolistic producer may have greater economies of scale compared to monopolistic competitor. In some industries, this could mean lower prices for consumers (such as in regulated, natural monopolies like utilities).

Conclusion: For the vast majority of the goods bought by the typical consumer, monopolistic competition is a preferable market structure to monopoly or perfect competition.

- In the case of homogeneous primary commodities (such as agricultural crops like corn or soybeans), perfect competition is preferable, since more competition means lower prices for products that consumers do not necessarily care about product variety for.
- In the case of large scale utilities and infrastructure goods, like a city's water supply or electricity grid, monopoly may be the preferred market structure, since the ability for a single firm to achieve economies of scale allows for the regulated firm to

- provide the service for a lower price to consumers.
- In the case of most consumer goods, such as restaurants, salons, apparel, consumer electronics, and so on, monopolistic competition is the preferred market structure, since product differentiation and non-price competition generally result in a better quality and more product variety for consumers, while competition and low entry barriers assure relatively low prices and a more efficient level of output than would result from monopoly.

Assumptions of the oligopoly market model

- Describe, using examples, the assumed characteristics of an oligopoly: the dominance of the industry by a small number of firms; the importance of interdependence; differentiated or homogeneous products; high barriers to entry.
- Explain why interdependence is responsible for the dilemma faced by oligopolistic firms - whether to compete or to collude.
- Explain how a concentration ratio may be used to identify an oligopoly.

The final market structure we will study lies between monopolistic competition and pure monopoly on the competitive spectrum. An **oligopoly** is an industry with a few large sellers, each with a substantial share of the total market demand.

Characteristic	Oligopoly
Number of Firms	A few large firms dominate the industry, each with a substantial share of total demand. There are few enough firms that in some cases, collusion is possible (when firm coordinate price and output decisions). Collusion can be: <ul style="list-style-type: none"> Open / formal Tacit / informal
Price making abilities of individual firms	A change in one firm's output has significant impact on the market price, firms are price-makers.
Type of product	Products can be identical (such as oil) or differentiated (such as Apple computers and PCs) Firms will likely use advertising to try and differentiate their products from competitors'
Entry barriers	There are significant barriers to entry, such as economies of scale, legal barriers, ownership of resources, etc...

Examples of oligopolistic markets

Oligopoly is a relatively common form of market structure. Many of the consumer goods and services we demand are provided by oligopolistic firms, including:

- Cell phone service providers: In most countries, consumers will have only a few choices for whom to buy their cell plan from. The providers all differentiate through options such as text messaging, data plans, call time, etc...
- Airplane manufacturers: Boeing and Airbus are the two dominant firms in the market for jumbo-jets. The firms differentiate through fuel efficiency of their craft,

- number of seats, and so on.
- Movie studios: Only six big Hollywood studios make over 90% of the movies that make it to the big screen.
 - Beer in the United States: Despite the fact that there are thousands of independent breweries in the US, only two large corporations produce 80% of the total beer supply. Both firms offer dozens, perhaps hundreds of varieties to try to differentiate their product from the competition
 - Petrol for cars: Automobile fuel is a product often sold by a handful (a dozen or so) of large firms. Fuels, unlike the other products above, are a homogeneous product, so firms differentiate through location, primarily.

Oligopolistic producers are interdependent upon one another. **Interdependence** refers to the state of two or more entities (firms, in this case) that must consider one another's actions or decision and when making their own actions or decision.

For example, consider the market for cell phone providers in the United States, which has only four dominant firms - AT&T, Verizon, T-Mobile, and Sprint. If T-Mobile begins offering customers unlimited high-speed data at only \$50 per month, while all other carriers cap data usage, the other three firms must consider the impact T-Mobile's decision will have on demand for their service.

- When a single firm in an oligopolistic market changes its price or its product, demand for other firms' products will be affected.
- A price cut by one seller will reduce demand for all others.
- A price increase by one seller will increase demand for all others.

Unlike in more competitive markets structures, oligopolies must respond to actions taken by their competitors lest demand for their product be negatively affected. This interdependence among the firms in oligopoly markets is an important characteristic that has implications for our analysis of oligopoly behavior.

One way to determine whether a market is oligopolistic is to consider the market's **concentration ratio**, which indicates the percentage of total sales controlled by the largest firms in a market. For example, consider the US cellular market again. The four dominant firms, AT&T, Verizon, T-Mobile, and Sprint, each enjoy the market shares indicated in the table below.

Firm	Market share
Verizon	35%
AT&T	32%
T-Mobile	17%
Sprint	14%

Other providers	2%
-----------------	----

The US cellular provider market's **four-firm concentration ratio** is **98%**. In other words, the four dominant firms account for 98% of total sales in the market.

The general rule is that if the four-firm concentration ratio of an industry is greater than 60%, the industry is an oligopoly. It doesn't matter how many firms actually compete in the market; if the four largest account for at least 60% of total sales, the market is oligopolistic.

Game theory

- Explain how game theory (the simple prisoner's dilemma) can illustrate strategic interdependence and the options available to oligopolies.

Because there are only a few large firms in oligopolistic markets, they often have a strong incentive to cooperate, rather than compete, with one another on output and pricing decisions.

To understand why collusion is so attractive to oligopolistic firms, it is useful to think of competition between them as a sort of game. For this, we will use a model of oligopoly behavior known as game theory.

Game Theory is the study of strategic decision making through the use of games

Consider the following example. Two cellular service companies, Firm A and Firm B, provide cell phone service in a country. These firm are trying to decide on the following:

- Whether to offer unlimited data to their customers (we will refer to this option as FREE), or
- Whether to charge customers based on data usage (we will refer to this option as PAY)

The profit of each firm depends not only on whether it offers free data, but also on whether its competitor offers free data. In this regard, the firms are highly interdependent on one another

The possible levels of profit Firms A and B can earn depending on their decision regarding data plans AND based on the competition's decision can be plotted in a table called a payoff matrix.

Study the payoff matrix below:

Payoff Matrix		Firm B	
		PAY	FREE
Firm A	PAY	10 , 10	5 , 20
	FREE	20 , 5	7 , 7

In this “game”:

- Each firm can either choose “PAY” or “FREE”
- The number on the left in each box is the possible level of economic profit (in millions of dollars) enjoyed by Firm A.
- The number in the right in each box is the possible profit earned by Firm B.
- Notice that each firm’s profit depends largely on what the competition chooses to do.

Determining the likely outcome of the game

Assume the firms do not collude. What will each firm most likely do? To determine the most likely outcome in the game below, consider the possible payoffs the firms face.

If Firm A chooses “PAY”

- And Firm B also chooses PAY, A will earn profits of \$10 million
- But if Firm A chooses FREE, B’s profits will fall to \$5 million

If Firm B chooses “FREE”

- And Firm A chooses PAY, Firm B will earn profits of \$20 million
- But if Firm A also chooses FREE, Firm B’s profits will be \$7 million.

Determining a dominant strategy:

A strategy is dominant if it results in a higher payoff regardless of what strategy the opponent chooses.

- In this game, both firms have a dominant strategy of choosing FREE.
- If Firm A chooses PAY, Firm B can do better by choosing FREE.
- If Firm A chooses FREE, Firm B can do better by choosing FREE.

Payoff Matrix		Firm B	
		PAY	FREE
Firm A	PAY	10, 10	5, 20
	FREE	20, 5	7, 7

Both firms can always do better by choosing to offer FREE data!

This game is known as the Prisoner's Dilemma. The firms in the game face a dilemma because:

- Both firms want to maximize their own profits, but...
- The rational thing to do is to offer FREE data, because the potential profits are so great!
 - \$20 million if the competitor chooses PAY, and
 - \$7 million if the competitor chooses FREE,
 - For a total possible payoff of \$27 million
- The possible payoffs for offering PAY are lower
 - \$10 million if the competitor offer PAY, and
 - \$5 million if the competitor offers FREE,
 - For a total possible payoff of \$15 million

When they act in their own rational self-interest, both firms end up earning less profits than if they had instead acted irrationally.

The dilemma is that, ultimately, the firms are likely to earn LESS total profits between them by offering FREE data than they would have earned if they had only chosen PAY data. This is because collusion was not possible.

Game theory teaches us that in oligopolistic markets:

- Firms are highly interdependent on one another and that...
- There is a good reason for firms to collude with one another, because
- Through collusion, firms can choose a strategy that maximizes total profits between them, however...
- Such an outcome (both firms choosing PAY in our game) is highly unstable, because both firms have a strong incentive to cheat.

Payoff Matrix		Firm B	
		PAY	FREE
Firm A	PAY	10 , 10	5 , 20
	FREE	20 , 5	7 , 7

Game theory in the real world

This model of oligopoly behavior can be used to analyze the behavior of firms in oligopolistic markets on several levels, including:

- Whether to set a high price or a low price,
- Whether to advertise or not,
- Whether to offer free customer service
- Whether to offer a 1 year warranty or a three year warranty,
- Whether to open a store in a certain location or not... and so on...

In each of these scenarios, the decision one oligopolist makes will impact not only its own level of profits, but also those of its close competitors.

Open/formal collusion

- Explain the term “collusion”, give examples, and state that it is usually (in most countries) illegal.
- Explain the term “cartel”.
- Explain that the primary goal of a cartel is to limit competition between member firms and to maximize joint profits as if the firms were collectively a monopoly.
- Explain the incentive of cartel members to cheat.
- Analyse the conditions that make cartel structures difficult to maintain.

Collusion is defined as the open or tacit cooperation between firms in an oligopolistic to set prices or agree on other strategies that often benefit the firms at the expense of consumers.

In order to formally collude, the firms in a particular industry may form an official organization through which price and output decisions are agreed upon. This is called a CARTEL

- Cartels are illegal in most industries in most countries, due to their anti-competitive nature
- The firms in a cartel will choose an output and price that a monopolist would choose
- The price consumers pay will be higher, the output lower (consumer surplus lower)
- Cartels tend to stifle innovation among firms and reduce both productive and allocative efficiency.
- Due to the prisoner's dilemma explained earlier (there is always an incentive to cheat in

a collusive oligopoly), cartel arrangements are often unstable and difficult to maintain. Once the majority of firms have agreed to a high price and reduced output, each individual firm has a strong incentive to increase its output to take advantage of the higher price in the market. If all firms do this, the market price will fall and the cartel will fail.

Examples of cartels:

- OPEC (Organization of Petroleum Exporting Countries)
- International sugar producers
- International coffee growers
- Drug cartels of Latin America.

Tacit/informal collusion

- Describe the term “tacit collusion”, including reference to price leadership by a dominant firm.

Since formal collusion is illegal in many countries, oligopolistic firms have devised way to collude informally (also called **tacit collusion**). The most common form of tacit collusion is Price Leadership:

- Price leadership: This is when the biggest firm in an industry sets a price and the smaller firms follow suit. If the price leader raises its price, the competitors will too. If it lowers price, smaller firms will follow.
- Usually a "dominant firm" (typically the largest in the industry) establishes the price and smaller firms follow.
- Prices tend to be "sticky" upwards, since firms are hesitant to raise prices and lose market share to rivals.
- However, prices are "slippery" downwards, which means if one firm lowers its prices, others will follow suit so they don't lose all their business.

When tacit agreements break down, firms may engage in **price wars**, in which they continually lower their prices and increase output in order to try and attract more customers than their rivals.

- This can cause sudden increases in output and decreases in price, temporarily approaching an efficient level.
- Once firms realize low prices hurt everyone, price leadership is usually restored, and prices rise once more.

Non-collusive oligopolies

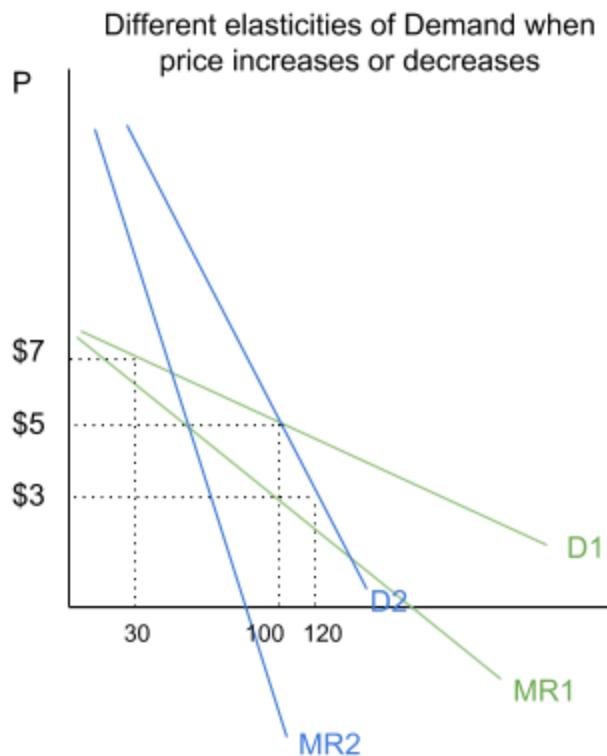
- Explain that the behaviour of firms in a non-collusive oligopoly is strategic in order to take account of possible actions by rivals.
- Explain, using a diagram, the existence of price rigidities, with reference to the kinked demand curve.
- Explain why non-price competition is common in oligopolistic markets, with reference to the risk of price wars.
- Describe, using examples, types of non-price competition.

What if collusion is not possible? Price and output decisions in oligopolies can be analyzed using a more traditional model of firm behavior, the demand curve.

Consider the market for hamburgers. Assume there are only two firms selling hamburgers, McDonald's (the Big Mac) and Burger King (the Whopper).

- The current price of both Big Macs and Whoppers is \$5.
- McD's is considering changing its price.
- If McD's lowers its price, it should assume that BK will also lower its price, because if they do not, they will lose many consumers to McD's.

The graph below shows two demand/MR curves, one showing what happens if McD's raises its price from \$5 (D₁), and one showing what happens if McD's lowers its price (D₂).



Observe from the graph:

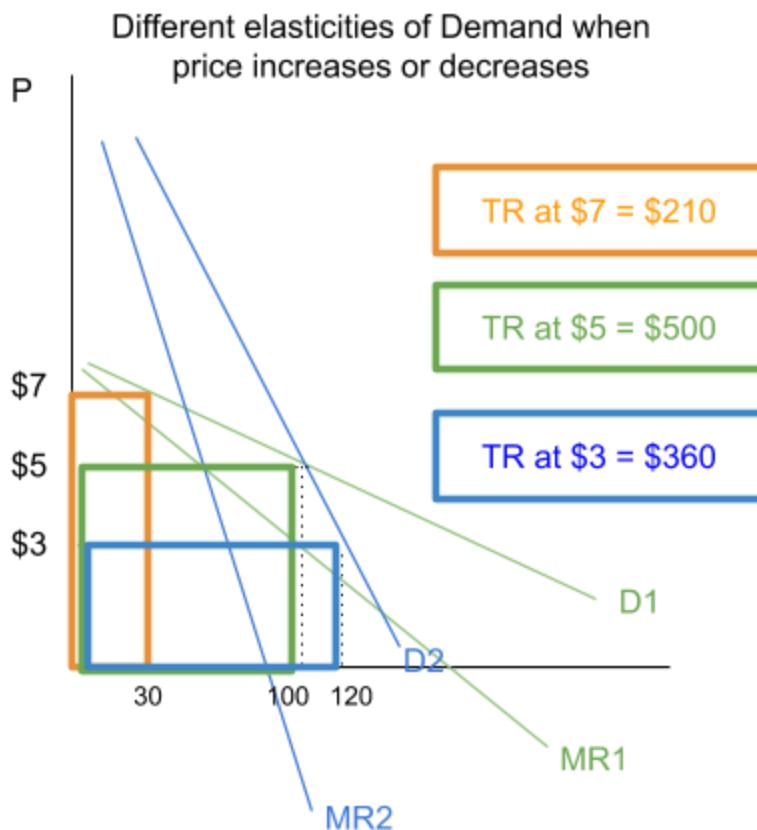
- Demand for Big Macs is highly inelastic below \$5 (D₂). Very few new customers will buy Big Macs, since the price of Whoppers will also fall. A 40% decrease in price (from \$5 to \$3) will only lead to a 20% increase in Qd (to 120).
- Demand for Big Macs is highly elastic above \$5. If McD's raises its price, BK will ignore the price increase, since BK knows lots of Big Mac consumers will switch over to Whoppers. A 20% increase in price (from \$5 to \$7) will lead to a 70% decrease in quantity (to 30).

The demand for an oligopolist's output in a non-collusive market is highly elastic above its

current price and highly inelastic below the current price.

This means that prices in oligopolistic markets tend to be highly stable, as firms rarely have the incentive to raise or lower price, since any change will likely cause a decrease in revenues.

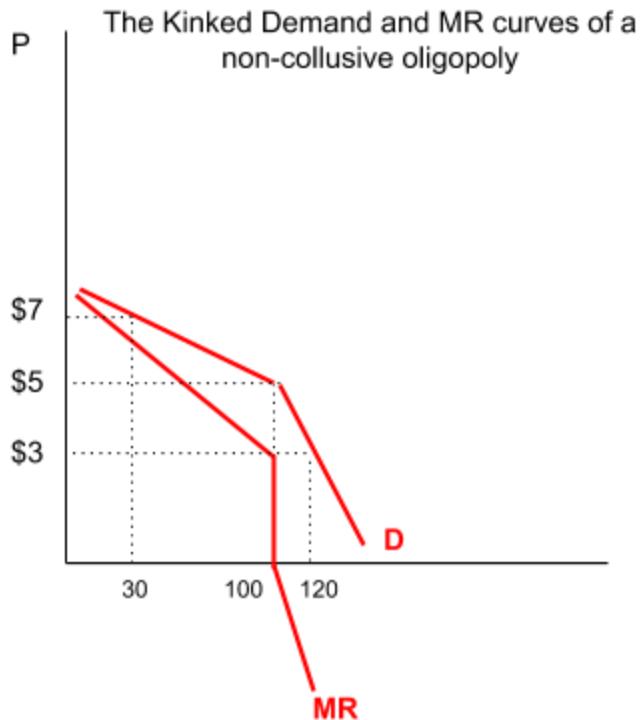
The graph below shows the impact of a change in the price of the Big Mac on McDonald's daily revenues assuming demand is highly elastic above \$5 and highly inelastic below \$5.



Observe from the graph:

- At \$5 McD's earns revenues of \$500.
- By raising its price to \$7, the firm's revenues fall to \$210.
- By lowering its price to \$3, the firm's revenues fall to \$360.
- McD's has little incentive to change its price, as doing so will only reduce the firm's total revenues.

Based on the analysis above, we can conclude that the demand for Big Macs, as seen by McDonald's is actually a **kinked demand curve**.

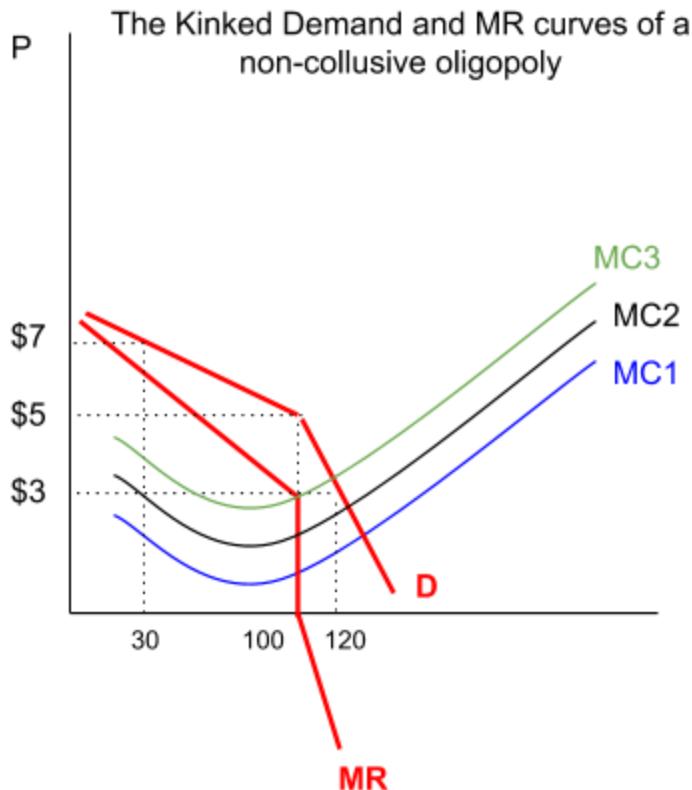


In the kinked demand and marginal revenue curves above we can see the following:

- Demand is highly elastic above the current price since price increases will be ignored and consumers will switch to other firms' products if price is raised.
- Demand is highly inelastic below the current price since price decreases will be matched by the competition and there will be very little increase in sales at the now lower price.
- Marginal revenue is negative at a quantity greater than 100 because the firm's total revenues will fall if it lowers its price.

The price in a non-collusive oligopolistic market tends to be very stable. Firms are unlikely to raise or lower prices since in either case total revenues will fall, possibly reducing profits.

Even as a firm's costs rise and fall, the firm is not likely to quickly change its level of output and price in a non-collusive oligopoly. Observe the graph below:



Assume due to rising beef prices, the marginal costs of Big Macs has risen from MC1 to MC3

- Following its profit maximization rule of producing where $MC=MR$, McD's should not change its price or quantity, even as the price of beef rises.
- Only if marginal cost rose higher than MC3 would McD's have to raise its price and reduce its output to maintain its profit maximizing level.
- Only if marginal cost fell lower than MC1 would McD's have to lower its price and increase its output to maintain profit maximization.

Prices and output are highly inflexible in a non-collusive oligopolistic market!

Since oligopolists try to avoid competing on price, **non-price competition** is a common characteristic of oligopolistic markets. Examples of non-price competition include:

- **Advertising:** Dominant firms in oligopolistic markets tend to spend more on advertising than firms in more or less competitive markets, with the goal of “educating” consumers about the attributes of their product that set them apart from the other dominant competitors in the market.
- **Product differentiation:** As in monopolistic competition, oligopolists tend to differentiate their products to set them apart from the competition in an effort to increase sales and revenues without the risk of engaging the competition in a price war.
- **Brand loyalty:** Oligopolistic sellers go to great effort to foster brand loyalty among

their consumers. Examples include “frequent flyer” programs among airlines, which reward customers for always choosing the same airline when booking travel, “customer loyalty cards” among grocery retailers which give customers discount and access to promotions, and even contracts that limit the ability of a customer to switch to a competitor’s products, such as the one or two year commitments cellular phone customers are required to enter into when choosing a provider.

Price discrimination

- Describe price discrimination as the practice of charging different prices to different consumer groups for the same product, where the price difference is not justified by differences in cost.
- Explain that price discrimination may only take place if all of the following conditions exist: the firm must possess some degree of market power; there must be groups of consumers with differing price elasticities of demand for the product; the firm must be able to separate groups to ensure that no resale of the product occurs.
- Draw a diagram to illustrate how a firm maximizes profit in third degree price discrimination, explaining why the higher price is set in the market with the relatively more inelastic demand.

In our study of price-making, imperfectly competitive firms, we have so far assumed that each firm charges a single price to all consumers for the product in question. However, all consumers do not always pay the same price for a particular good.

When firms charge different prices to different consumers for the same product they are practicing what is known as **price discrimination**.

In order for a firm to be able to price discriminate, the following conditions must be met:

1. The firm must have some monopoly power: A perfect competitor could not possibly charge different prices to different consumers, because there are hundreds of other firms selling the same good for the low market price.
2. Market Segregation: In order to price discriminate, the seller must be able to determine who is willing to pay what. The firm must segregate the market by age, gender, race, nationality, income level, or some other method that distinguishes between consumers willing to pay more for a good and those willing to pay less.
3. No Resale: If a buyer who paid a low price is able to sell the product to someone who the seller wants to charge a high price to, the seller's monopoly power is undermined and it becomes difficult to price discriminate. So it must be difficult or impossible for buyers to resell the product to one another.

Price discrimination comes in many forms, and we distinguish between different degrees of price discrimination.

- **First degree price discrimination** – by individual consumer: This is the most difficult type of price discrimination for firms to practice. It requires the firm to determine exactly what each consumer is willing and able to pay for the product, and charges each consumer that price. This type leads to the greatest profits for the firm,

- but leaves consumers with no consumer surplus. Sometimes referred to as perfect price discrimination.
- **Second degree price discrimination** - by quantity: A more common form of price discrimination in which the firm charges lower per-unit price to consumers who “buy in bulk”. Consider a pack of toilet paper rolls with four rolls in it compared to a package with 24 rolls in it. Usually, if you buy the larger package, you will pay considerably less per roll. This is a form of price discrimination which charges higher prices to people who are not willing to buy in bulk.
 - **Third degree price discrimination** – by consumer group: Another common form of price discrimination; consumers may pay more or less for a good depending on their age, their gender, when they buy, the passport they carry, etc... Consider movie theater tickets (age), airline tickets (when you buy), haircuts (gender) and admission to museums or national parks in some countries (nationality).

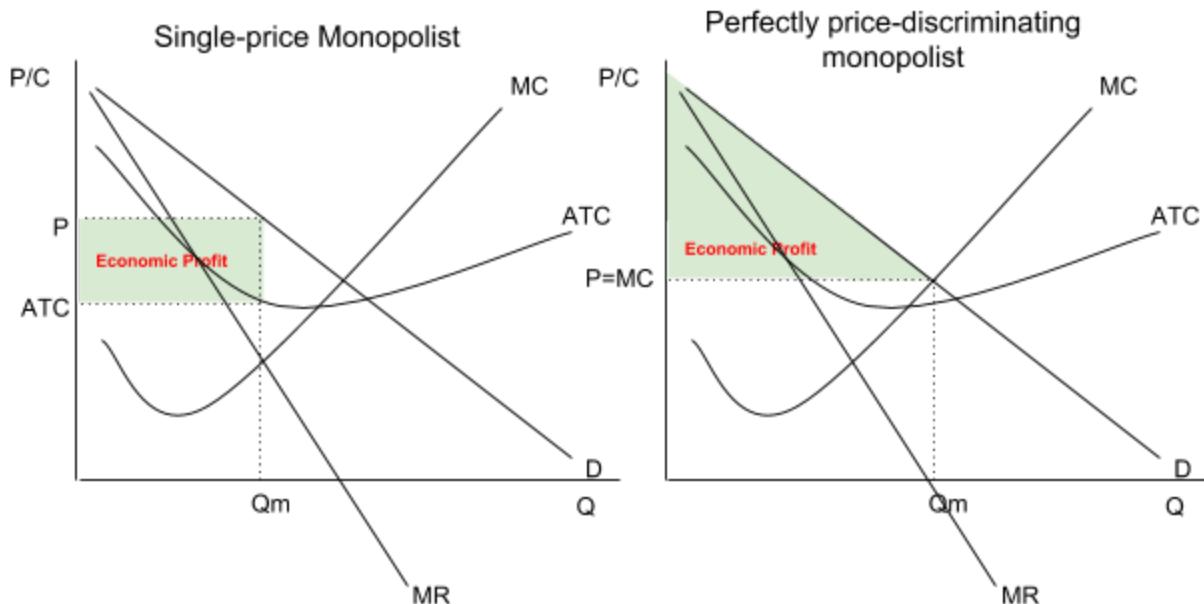
There are countless examples of firms price discriminating. Here are a few:

- Movie theaters: Charge different prices based on age. Seniors and youth pay less since they tend to be more price sensitive.
- Gas stations: Gas stations will charge different prices in different neighborhoods based on relative demand and location.
- Grocery stores: Offer coupons to price sensitive consumers (people whose demand is inelastic won't bother to cut coupons, thus will pay more for the same products as price sensitive consumers who take the time to collect coupons).
- Quantity discounts: Grocery stores give discounts for bulk purchases by customers who are price sensitive (think “buy one gallon of milk, get a second gallon free”... the family of six is price sensitive and is likely to pay less per gallon than the dual income couple with no kids who would never buy two gallons of milk).
- Dell Computers: Dell price discriminates based on customer answers to questions during the online shopping process. Dell charges higher prices to large business and government agencies than to households and small businesses for the exact same product!
- Hotel room rates: Some hotels will charge less for customers who bother to ask about special room rates than to those who don't even bother to ask.
- Telephone plans: Some customers who ask their provider for special rates will find it incredibly easy to get better calling rates than if they don't bother to ask.
- Damaged goods discounts: When a company creates and sells two products that are essentially identical except one has fewer features and costs significantly less to capture more price-sensitive consumers.
- Book publishers: Some paperbacks cost more to manufacture but sell to consumers for significantly less than hard covers. Price sensitive consumers will buy the paperback while those with inelastic demand will pay more for the hard cover.
- Airline ticket prices: Weekend stopover discounts for leisure travelers mean business people, whose demand for flights is highly inelastic, but who will rarely stay over a weekend, pay far more for a round-trip ticket that departs and returns during the week.

The effects of price discrimination can be shown graphically, which allows us to determine whom benefits, which suffers, and whether it increases or decreases overall welfare and efficiency.

First degree price discrimination

The graphs below compare a single price monopolist and a perfectly price discriminating firm.



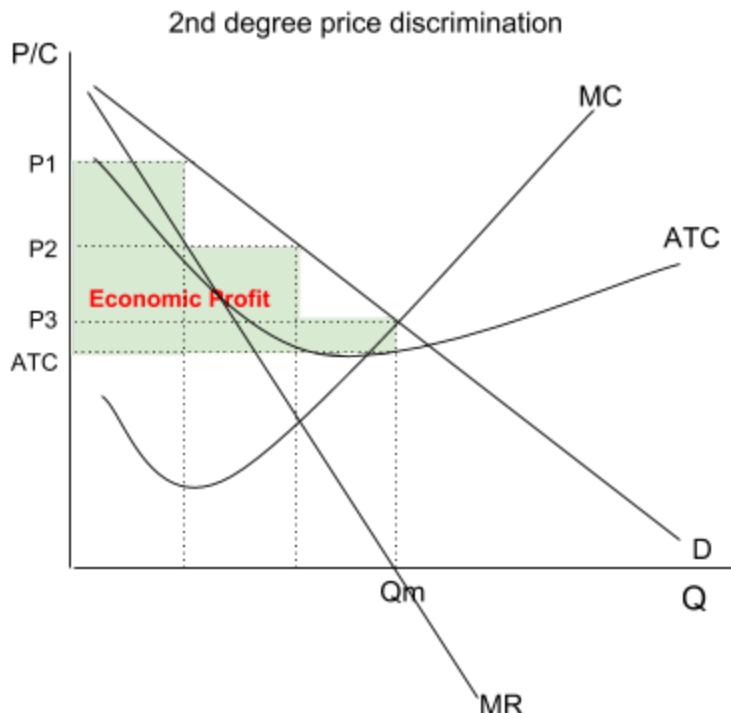
Examining the graphs above, we can make the following observations about price discrimination:

- The price discriminating firm earns a greater level of economic profits than the single-price firm. The shaded triangle on the right is bigger than the shaded rectangle on the left
- More output is produced and sold due to price discrimination: Q_{pd} is greater than Q_m
- Consumer surplus is reduced (or eliminated in the case of perfect price discrimination). When every consumer pays exactly what she is willing to pay, no one has any “extra” happiness when buying the product.
- Allocative efficiency is improved! The higher level of output will be closer to (or equal to in the case of perfect price discrimination) the $P=MC$ level. The firm will continue to sell right up to the point the last price it charged is equal to the firm's marginal cost.
- More efficient allocation of resources: Despite the fall in consumer surplus, overall welfare is actually improved by price discrimination. More people can afford the product than under a single price seller.

Second degree price discrimination

In 2nd degree price discrimination firms do not know much about the characteristics of different consumers and thus are unable to segregate the market based on consumer

incomes or other factors. Rather, firms charge different prices based on the quantity consumers buy. Examine the graph below.



Observe from the graph:

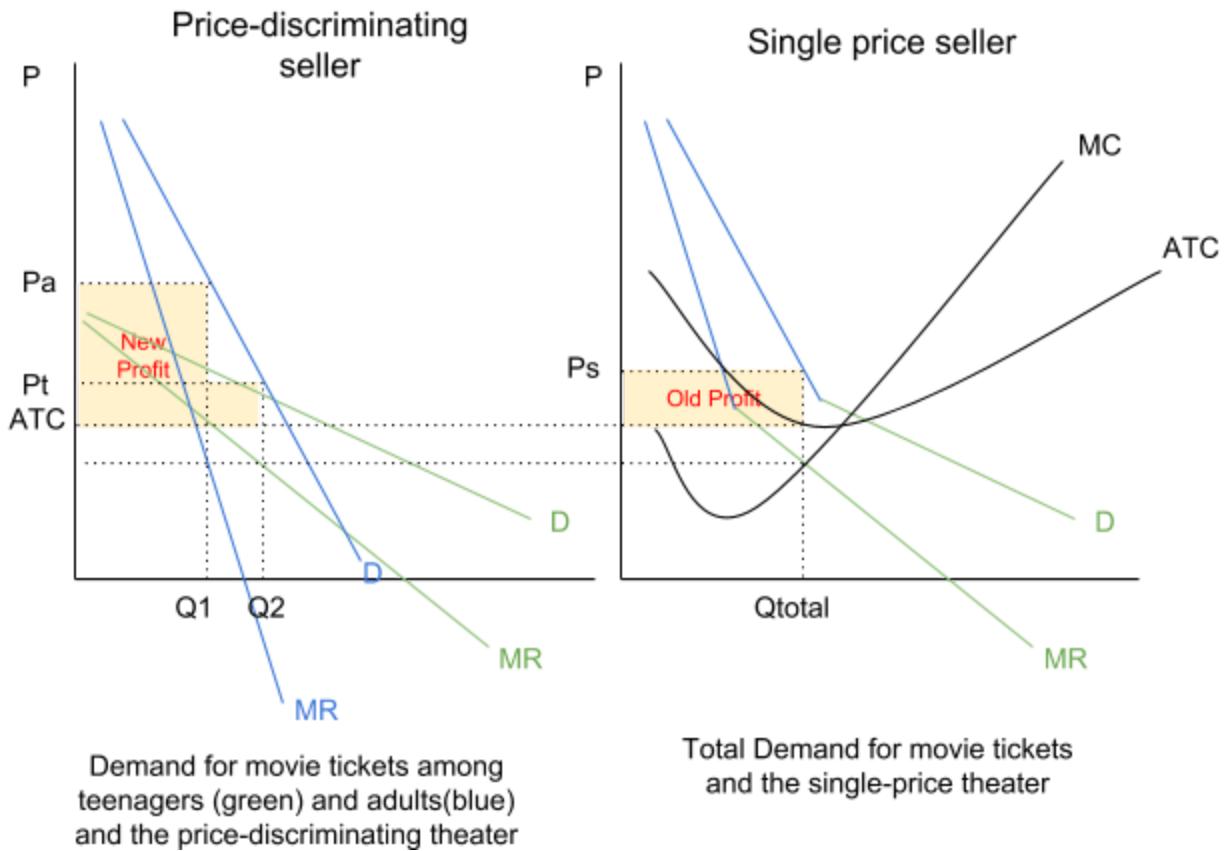
- Consumers who buy in lower quantities pay higher price of P_1 , while consumers who buy in bulk pay lower prices of P_2 and P_3 .
- By charging different prices the firm is able to extract more consumer surplus and earn more economic profit.
- Total output is higher under 2nd degree price discrimination than under a single-price seller, and total welfare is thus greater.
- Consumers who would have been unable to buy the good under the single price can now afford the good.

Bulk discounts are commonly used by consumer goods retailers: consider large packages of toilet paper versus smaller packages, volume discounts on cleaning or cooking supplies, and other deals offered by “big box” retailers like Walmart, Costco, Tesco, Carrefour and others.

Third degree price discrimination

Firms that are able to segregate their consumers based on characteristics such as income level, gender, nationality, age, etc... may employ 3rd degree price discrimination. The idea is to charge consumers whose demand is greater and less elastic higher prices while charging consumers whose demand is lower and more elastic lower prices.

The graphs below show the demand and marginal revenues for movie theater tickets on a Friday night among teenagers (in green) and adults (in blue) and the total demand of both groups in the graph on the right.



Observe from the graphs:

- Teenagers' demand (in green) is more elastic than demand from adults (in blue).
- If the movie theater had to set a single price (P_s) for movie tickets it would base it on the combined demands of the two subgroups and earn an area of profit like that shown in the graph on the right.
- By carrying over the firm's MC from the graph on the right to the graph with both groups' demand on the left, the firm can set prices above where $MC=MR$ for teenagers (P_t) and a higher price above where $MC=MR$ for adults (P_a)
- By charging higher prices to adults and lower prices to teenagers, however, the theater is able to extract more consumer surplus and convert it to economic profit (as in the graph on the left).

Charging higher prices to consumers with relatively inelastic demand (adults who buy movie tickets) and lower prices to consumers with relatively elastic demand (teenagers), sellers are able to both earn more profits and sell to more consumers than they would if only a single price were set.

Other examples of third degree price discrimination include: airplane tickets, museum admission, cell phone plans, hair salons, night clubs (ladies night), and so on.

2.1 The Overall Level of Economic Activity

The circular flow of income model

- Describe, using a diagram, the circular flow of income between households and firms in a closed economy with no government.
- Identify the four factors of production and their respective payments (rent, wages, interest and profit) and explain that these constitute the income flow in the model.
- Outline that the income flow is numerically equivalent to the expenditure flow and the value of output flow.
- Describe, using a diagram, the circular flow of income in an open economy with government and financial markets, referring to leakages/ withdrawals (savings, taxes and import expenditure) and injections (investment, government expenditure and export revenue).
- Explain how the size of the circular flow will change depending on the relative size of injections and leakages.

Introduction to macroeconomics

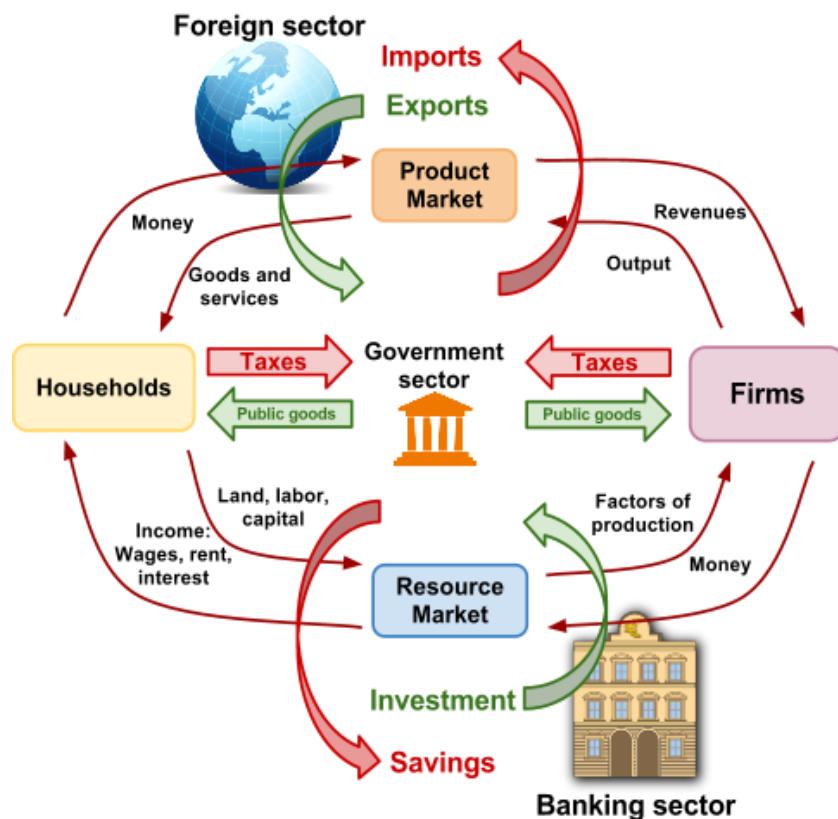
There are several key concepts from Microeconomics that are similar to some of the topics that will be covered in Macroeconomics. The table below shows several of the Micro concepts and their Macro equivalents.

Micro Concept	Macro Concept	Key Terms in Macroeconomics
Market	National Economy	Examines all the economic activity taking place in a country
Demand	Aggregate Demand (AD)	The total demand for a nation's output of goods and services
Supply	Aggregate Supply	The total supply of goods and services by all the industries of a country
Price	Average Price Level	An index of the average prices of goods and services over time
Quantity	National Output	Total output of all the industries of a country; also known as the gross domestic product (GDP), or national output / income
Decrease in Demand	Recession	A fall in total output resulting from a decrease in AD
Increase in Demand	Inflation	An increase in the average price level resulting from an increase in AD
Decrease in Supply	Supply Shock	An increase in the price level and decrease in output from a fall in AS

Increase in Supply	Economic Growth	An increase in national output resulting from an increase in AS
--------------------	-----------------	---

The simplest model representing the economic activity in a country is **the circular flow model**, which the flows of money between all the agents involved in the production and consumption of goods and services in a nation:

- **Households** provide factors of production to firms in the factor market and demand goods and services in the product market.
- **Firms** produce goods and services for households in the product market using resources bought from households in the factor market.
- The **government** collects taxes from households and firms, from which it provides goods and services (such as infrastructure, defense, and education) to households.
- The **foreign sector** includes the demand for goods and services produced by a nation in foreign countries and the demand for foreign goods by the nation's households and firms.
- The **banking sector** provides households and firms with loans and offers a place to save and invest money.



Leakages and injections a nation's circular flow

In the circular flow model above there are arrows indicating leakages from and injections to the circular flow. Money leaked from a nation's economy is subtracted from the national income; money injected into the economy is added to the national income.

- **Leakages:** Taxes paid to the government, spending on imports from abroad, and money saved in banks are all considered leakages from the circular flow of income. Any income earned but NOT spent on goods and services does not contribute to the nation's total output, and is therefore leaked from the nation's economy. However, these three leakages allow for the three following injections.
- **Injections:** Government spending, export revenues and investments are all enabled by the three leakages above.
 - Because households and firms pay taxes, government has money to provide the nation with valuable infrastructure, education, defense, support for health care and so on, all public or quasi-public goods that would be under-provided by the free market. These contribute to national output and are thus injections into the circular flow.
 - Because domestic households buy imports, foreigners have access to the money the need to buy the nation's exports. The spending by foreigners on domestically produced goods contributes to national output and is therefore an injection.
 - Because households save some percentage of their income, capital is available for others to borrow and spend. Spending on capital goods by firms or on homes by households (both considered investments) contributes to the nation's output and is thus an injection into the circular flow.

Measures of economic activity: gross domestic product (GDP), and gross national product (GNP) or gross national income (GNI)

- Distinguish between GDP and GNP/GNI as measures of economic activity.
- Distinguish between the nominal value of GDP and GNP/GNI and the real value of GDP and GNP/GNI.
- Distinguish between total GDP and GNP/GNI and per capita GDP and GNP/GNI.
- Examine the output approach, the income approach and the expenditure approach when measuring national income.
- Evaluate the use of national income statistics, including their use for making comparisons over time, their use for making comparisons between countries and their use for making conclusions about standards of living.

HL only objectives:

- Calculate nominal GDP from sets of national income data, using the expenditure approach.
- Calculate GNP/GNI from data
- Calculate real GDP, using a price deflator.

Gross Domestic Product (GDP) measures the total flow of income between households, firms, the government, banks, and foreign countries. GDP is found by adding up the total value of the output produced within a country's borders in a particular period of time or by adding up all the income earned by the country's households.

GDP measures national output and national income, the idea being that the value of a nation's output is equal to the value of all the income earned by the households in a nation in a period of time.

Three ways of measuring GDP

Looking closely at the circular flow model, we can see that there is a relationship between the amount of income earned, the expenditures made, and the total output. The economic activity of a nation can, in fact, be measured using any of these three methods.

The income approach measures GDP by recording the income of household in the resource (factor) market side of the circular flow of income. Income includes payments households receive in the resource market in exchange for providing firms with the factors of production, including the total sum of each of the following earned by a nation's households in a year: Wages for labor, Interest for capital, Rent for land and Profits for entrepreneurship.

$$\text{GDP} = W + I + R + P$$

The value added approach measures the value of the total output produced in the different sectors of the economy. When the total output of every sector of the nation's economy is summed, total output is found.

$$\text{GDP} = \text{Outputs of the primary sector} + \text{the secondary sector} + \text{the tertiary sector}$$

The expenditure approach counts the total spending on final new goods and services in a given year. "Final" goods are ready for consumption and do not include goods that will be input goods or are raw materials for other production. This approach distinguishes between four types of spending on a nation's output. These include household consumption (C), investment in capital by firms (I), government spending (G) and exports (X) - imports (M) ($X - M$).

$$\text{GDP} = C + I + G + (X - M)$$

The expenditure approach to measuring GDP is the most commonly used in AP Macroeconomics. It measures the total spending on a nation's output by households, firms, the government and foreigners. The four types of spending are outlined below:

Household Consumption (C):

The purchase by households of all goods and services, including:

- Non-durables: bread, milk, toothpaste, t-shirts, socks, toys, etc...
- Durables: TVs, computers, cars, refrigerators, etc...
- Services: dentist visits, haircuts, taxi rides, accountants, lawyers, etc...

Gross Private Domestic Investment- (Ig)

- All final purchases of machinery, equipment, and tools by businesses.

<ul style="list-style-type: none"> • All construction (including residential). • Changes in business inventories <ul style="list-style-type: none"> ➢ If total output exceeds current sales, inventories build up. ➢ If businesses are able to sell more than they currently produce, this entry will be a negative number.
Government Purchases (of consumption goods and capital goods) - (G)
<ul style="list-style-type: none"> • Includes spending by all levels of government (federal, state and local). • Includes all direct purchases of resources (labor in particular). • This entry excludes transfer payments since these outlays do not reflect current production.

Net Exports- (Xn):

- All spending on goods produced in the U.S. must be included in GDP, whether the purchase is made here or abroad.
- Often goods purchased and measured in the U.S. are produced elsewhere (Imports).
- Therefore, net exports, (Xn) is the difference: (exports - imports) and can be either a positive or negative number depending on which is the larger amount.

GDP and GDP per capita

A nation's real GDP tells us the actual value of its output in a particular year, adjusted for any changes in the price level between that year and an earlier base year. However, real GDP does not tell us whether a nation is rich or poor. Consider the tables below:

Countries with largest GDP (and per-capita rank)	Total GDP, (trillions \$)	Country with largest per capita GDP (and total GDP rank)	Per-capita GDP (\$ in 2009)
1. United States (6)	14.2	1. Luxembourg (68)	105,350
2. Japan (14)	5.0	2. Norway (24)	79,089
3. China (86)	4.9	3. Denmark (29)	55,992
4. Germany (13)	3.3	4. Ireland (37)	51,049
5. France (12)	2.6	5. Netherlands (16)	47,917

GDP is a nation's total income. Because it does not account for the size of a country's population, **per capita GDP**, which measures the total GDP of a nation divided by the total population, provides a more accurate measure of how rich a nation is. Notice that none of the richest nations (on the right) are even in the top five for total GDP

What is included in GDP?

GDP measures the value of the final output of goods and services in a nation in a year. But there are some economic transactions which are not included in GDP.

GDP includes:

- GDP includes only final products and services
- GDP is the value of what has been produced within the borders of a nation over one year, not what was actually sold.

GDP Excludes "non production transactions":

- Purely financial transactions are excluded.
 - Public transfer payments, like social security or cash welfare benefits.
 - Private transfer payments, like student allowances or alimony payments.
 - The sale of stocks and bonds represent a transfer of existing assets (However, the brokers' fees are included for services rendered.)
- Second hand sales: If I buy a used car in 2013, that sale does not count towards 2013 GDP, because the car was not made in 2008! The price of the car was originally included in the year's GDP when it was produced.
- Non-market transactions: Economic activity that takes place in the informal sector or in the black market is not included in GDP. For example:
 - Volunteer work
 - Illegal activities (drug sales, crime, prostitution, etc)
 - Informal employment (babysitting, lawn mowing, etc)

GDP excludes environmental impacts of economic activity

- The negative environmental impact of economic transactions are not represented in the official GDP figure
- If production in an nation harms the environment or human health, GDP may overstate the true well-being of the nation's people

GDP excludes information about income distribution

- A high per person GDP does not mean everyone in a nation is well off
- High levels of inequality could mean that even in a nation with high average income, there are large portions of the population living in poverty

GDP excludes information about the composition of a nation's output

- The GDP figure does not indicate what is being produced in a nation
- A high GDP could mean a country is producing a lot of weapons or other goods that do not contribute to people's well-being
- A high GDP could mean a country is producing primarily capital goods, and not consumer goods, which may mean a lower standard of living for the nation's households.

Why is GDP important?

GDP is considered by economists to be the most important measure of economic activity in nations for several reasons:

- It tells us something about the relative size of different countries' economies
- It is a monetary measure, so it tells us how much income a country earns in a year (assuming everything that is produced is sold).
- When we divide GDP by the population, we get GDP per capita, which tells us how many goods and services the average person consumes in a country.
- When real GDP grows more than the population, that tells us that people on

- average, have more stuff than they did before.
- If you believe that having more stuff makes people better off, then GDP per capita tells us how well off people in society are.

Shortcomings of GDP

Because GDP excludes certain transactions, it can underestimate the true level of economic activity in a nation. Countries with large informal sectors or black markets could have much higher levels of income than the official GDP figure indicate.

Additionally, GDP excludes information about income distribution and the composition of the nation's output. All this is to say that GDP does NOT tell us how happy or well-off the people in a nation are; it is ONLY a measure of the monetary value of the goods and services produced in a nation. A country with a growing GDP or with a relatively high GDP per capita may be experiencing rising living standards or may have a higher quality of life than a country with a slower rate of growth or lower GDP, but not necessarily.

The shortcomings of GDP as a measure of well-being are summarized as follows:

- GDP ignores all social aspects of human life, such as income distribution, access to health care and education, life expectancy, gender equality, religious freedom, human rights and so on.
- Certain important work is left out of accounting (homemakers, labor of carpenters who make own homes because GDP measures only the MARKET VALUE of output. GDP therefore is understated.
- GDP does not reflect that people in most countries work fewer hours than in past years (in 1900 the average work week in the industrialized world was 53 hours, today it is around 40)
- Does not reflect improved product quality
- Does not include the underground economy
- GDP does not put a market value/cost on the environment. Higher GDP may be accompanied by negative externalities, which are NOT subtracted from GDP.
- GDP does not tell us if the best combo of goods and services are produced; a machine gun and textbooks are assigned equal weight.
- Nor does it measure how GDP is distributed among the population
- GDP does not measure the total well being, happiness, a reduction of crime or better relationships with society, with other countries, etc...

Alternatives to GDP

While gross domestic product is the primary measure of a nation's output in a particular year, economists have developed alternative measures of output which are sometimes referred to instead of GDP. These include:

Gross National Product (GNP) or Gross National Income (GNI) Measures the total value of output produced in a year by the factors of production provided by a nation.

- Differs from GDP in that it includes output produced abroad by domestically owned factories, but subtracts output produced domestically by foreign owned factories.

- Does not offer as accurate a measure of the actual economic activity within a nation as GDP does, and is therefore not considered as useful as GDP for measuring output of a nation.

Green GDP is an under-used measure of economic activity which subtracts from real GDP the losses to the environment and biodiversity resulting from economic growth.

- Places a monetary value on environmental degradation and subtracts this from the nation's GDP
- It is a measure preferred by environmentalists who believe that economic growth overstates increases in people's well-being due to the fact that it ignores the externalities that accompany growth.

Nominal GDP and Real GDP

Nominal GDP measures the value of a nation's output produced in a year expressed in the prices for that year. But if the average price level of a nation's output increases in a year, the nominal GDP would increase even if the actual amount of output does not change, since everything will appear more expensive at higher prices.

Real GDP measures the value of a nation's output in prices from a base year. By doing so, changes in the price level are ignored and the GDP figure reflects only whether actual output has increased or decreased over time. To determine the change in the real GDP.

- In the case of the price level increasing (inflation): real GDP will be lower than the nominal GDP
- In the case of the price level decreasing (deflation): real GDP will be higher than the nominal GDP

Converting Nominal GDP to Real GDP (HL only)

Study the output and price data for a hypothetical country in 2018 and 2019 below

Output in 2018	Quantity produced in 2018	Price in 2018	Total value of output 2018
Cheese	10	2	20
Chocolate	20	2	40
Watches	5	10	50
Nominal GDP in 2018:			110

Output in 2019	Quantity produced in	Price in 2019	Total value of output 2019

	2019		
Cheese	12	2.50	25
Chocolate	25	3	75
Watches	5	11	55
Nominal GDP in 2019:			160

To adjust a nation's nominal GDP in one year to its real GDP, we must measure the value of output using prices from a base year.

If we want to know the 2019 real GDP with 2018 as a base year, we must find the value of 2019's output in 2018 prices.

- 12 cheeses at \$2 = \$24
- 25 chocolates at \$2 = \$50
- 5 watches at \$10 = \$50
- 2019 real GDP = **\$124**

The GDP Deflator price index (HL only)

With the nominal and real GDP values for a given year, the GDP deflator price index can be calculated. The **GDP deflator** is a price index of everything produced in a nation in a period of time (unlike the consumer price index, which only includes the goods purchased by the typical household).

$$\text{GDP deflator price index} = \frac{\text{nominal GDP}}{\text{real GDP}} \times 100$$

2019's GDP deflator price index can be calculated:

$$\text{GDP deflator for 2019} = 160 \div 124 = 1.29 \times 100 = \mathbf{129}$$

Prices in 2019 were 129% of 2018's prices. In other words, the average price of the country's output grew by 29% between the two years.

Calculating real GDP using the GDP deflator (HL only)

The GDP deflator price index can be used to adjust a nation's nominal GDP for changes in the price level. The deflator is an indicator of how much prices have changed between two years.

- For a base year, the deflator always equals 100, since the real GDP = nominal GDP
- If, in a later year, the index is 110, this means that prices have risen by 10% between those years. If it is 120, prices have risen by 20%. If it is 95, then price fell by 5%, and so on...

To calculate real GDP the nominal GDP must be “deflated” by dividing it by the deflator

price index, then multiplying the result by 100.

$$\text{Real GDP} = \frac{\text{Nominal GDP}}{\text{GDP deflator price index}} \times 100$$

Consider the table below, showing nominal and real GDP data for the United States:

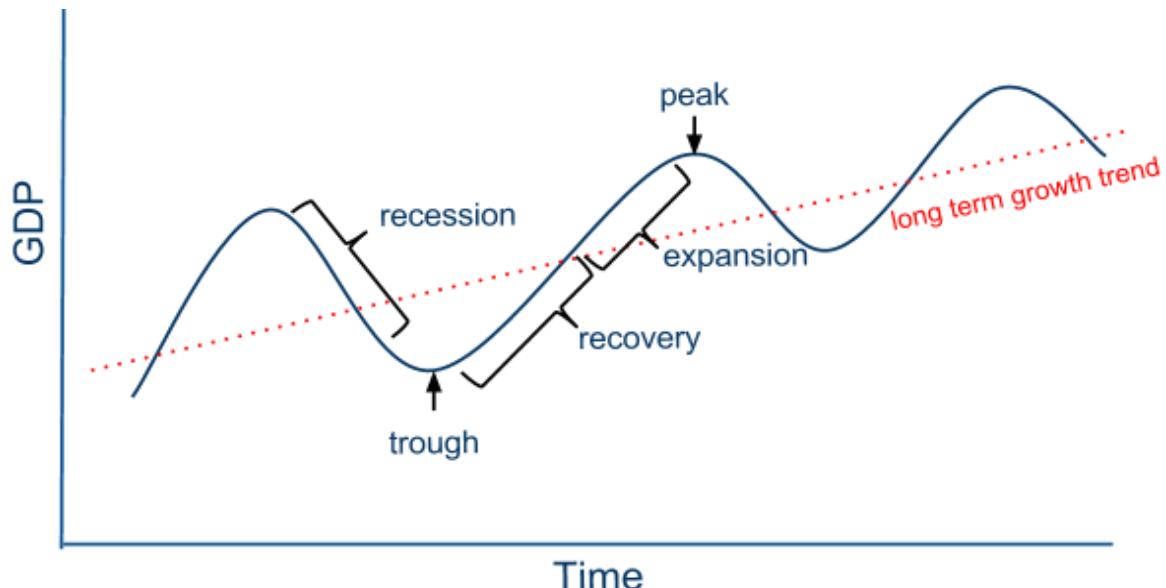
Year	Nominal GDP (billions of \$)	GDP Deflator	Real GDP (billions of \$)
2005	12,638.4	100	$\frac{12,638.4}{100} \times 100 =$ 12,638.4
2006	13,398.9	103.25	$\frac{13,398.9}{103.25} \times 100 =$ 12,976.2
2007	14,061.8	106.29	$\frac{14,061.8}{106.29} \times 100 =$ 13,228.9
2008	14,369.1	108.61	$\frac{14,369.1}{108.61} \times 100 =$ 13,228.8
2009	14,119.0	109.61	$\frac{14,119}{109.61} \times 100 =$ 12,880.6

Notice that for each of the years from 2006 on, real GDP was lower than nominal because the deflator increased each year, indicating that there was inflation; therefore, nominal GDP would have over-stated the changes in real output from year to year.

The Business Cycle

- Explain, using a business cycle diagram, that economies typically tend to go through a cyclical pattern characterized by the phases of the business cycle.
- Explain the long-term growth trend in the business cycle diagram as the potential output of the economy.
- Distinguish between a decrease in GDP and a decrease in GDP growth.

Changes in a nation's GDP over time can be illustrated in a simple economic model known as the **business cycle**. The business cycle model shows how a nation's GDP fluctuates over time, and typically demonstrates different phases.

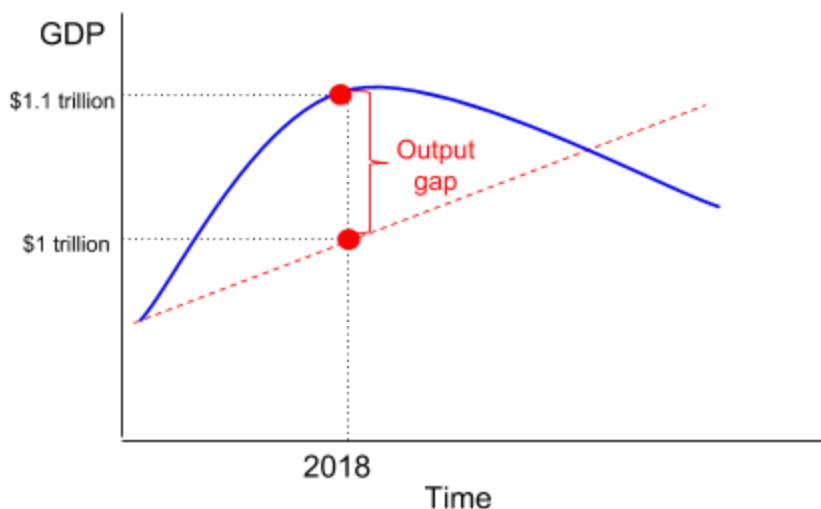


Phases of the business cycle:

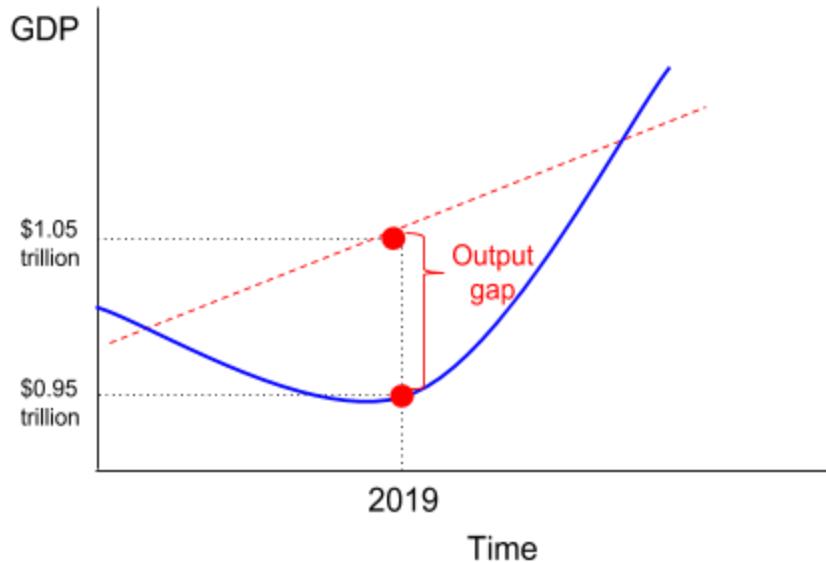
- **Expansion:** Output is increasing over time
- **Peak:** Output stops increasing and starts decreasing
- **Recession:** Output decreases over time
- **Trough:** Output stops decreasing and starts increasing

Output gaps in the business cycle

When an economy is producing above its long-run trend line, it is experiencing a **positive output gap**. When it is producing below the long-run trend line, it is experiencing a **negative output gap**.



In the graph above, the country is producing \$1.1 trillion of output in 2018, which is more than it would be producing if output had increased at its long-run average growth rate. However, output increased more than the long-run average and as a result the country is over-producing; it will soon pass a peak in its business cycles and enter a recession.



In 2019 the country has entered the recession phase and is experiencing a negative output gap of \$0.1 trillion. Potential output is \$1.05 trillion, but due to recession in the previous year its actual output is now below what it would be producing if the economy had grown at its long-run rate of GDP growth from the previous year.

Potential output is the GDP a country would experience if output increased at a steady rate over the long run and if factors of production are fully employed. When there is a positive output gap, actual output is greater than potential output; when there is a negative output gap, actual output is less than the potential output.

Long-run economic growth in the business cycle

Notice from the business cycle model that economic growth occurs over time, but not always at a steady rate. The red dashed lines in the graphs above represent the increase in the nation's potential output over time, or **long-run economic growth**. Of course, each economy's business cycle will look unique, but most economies will experience the types of fluctuations the model shows.

Possible causes of the business cycle

There are several theories regarding WHY countries experience economic fluctuations in the short run.

- Major innovations may trigger new investment and/or consumption spending.
- Changes in productivity may be a related cause.
- Most agree that the level of aggregate spending is important, especially changes in the purchase of capital goods and consumer durables.
- Cyclical fluctuations: Durable goods output is more unstable than non-durables and services because spending on latter usually cannot be postponed.

Decrease in GDP versus a decrease in GDP growth rate

- The growth rate of an economy refers to the percentage change in GDP between two periods of time. When an economy is approaching a peak in its business cycle, the rate of growth has begun to fall.
- When a recession begins, the actual output of an economy decreases. This means the growth rate has become negative.

2.2 Aggregate Demand and Aggregate Supply

The aggregate demand curve

- Distinguish between the microeconomic concept of demand for a product and the macroeconomic concept of aggregate demand.
- Construct an aggregate demand curve.
- Explain why the AD curve has a negative slope.

Introduction to Aggregate Demand

In microeconomics, the primary model used to show the interactions of buyers and sellers in market is the supply and demand model. In macroeconomics, economists are interested in total supply and total demand. The model we will use to examine the interactions of ALL the buyers and ALL the sellers in a nation's economy is the **aggregate demand and aggregate supply model**.

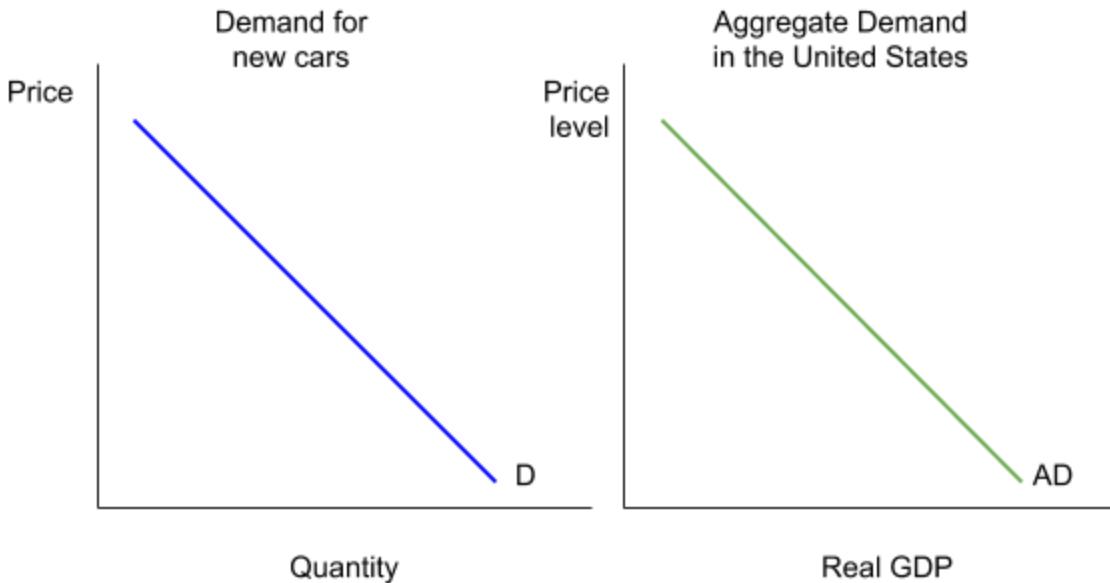
Aggregate demand (AD) describes the quantity of a nation's output demanded at a range of price levels by households, firms, the government, and the rest of the world.

Similarities between aggregate demand and demand:

- Both curves slope downwards, showing an inverse relationship between how much is demanded and prices
- There are 'non-price determinants' of both demand and aggregate demand. Changes in these factor will cause the curves to shift
- A decrease in both causes employment and output to fall. A fall in demand will cause output and employment in a particular industry to decrease; a fall in aggregate demand will cause output and employment in an entire country to decrease.
- An increase in both causes prices to rise. A rise in demand will cause the price of a particular good to increase; a rise in aggregate demand causes the average price level in an entire nation to increase (inflation).

The Aggregate Demand curve

The aggregate demand curve looks strikingly similar to a demand curve for a particular product. Let's examine the differences.



The graph on the left shows the demand for new cars in the United States. The graph on the right shows the aggregate demand for all goods and services produced in the United States. Notice the following:

- The AD curve shows the quantity demanded of all the goods and services produced in the US (its real GDP)
- The demand curve shows the quantity demanded in a single industry
- The AD curve shows the quantity of output demanded against the price level in the United States, which is measured using a price index such as the Consumer Price Index (CPI) or the GDP deflator price index.
- The demand curve shows the quantity of cars demanded against the price of cars
- Aggregate demand represents all industries in the United States.
- Demand represents a single industry.

Why does the AD curve have a negative slope?

There are three rationales for the inverse relationship between the price level in a nation and the quantity of the nation's output demanded by households, firms, the government, and the rest of the world:

- **The wealth effect:** Higher price levels reduce the purchasing power or the real value of the nation's households' wealth and savings. The public feels poorer at higher price levels, thus demand a lower quantity of the nation's output when price levels are high. At lower price levels, people feel wealthier and thus demand more of a nation's goods and services. (This is similar to the income effect which explains the downward sloping demand curve).
- **The interest rate effect:** In response to a rise in the price level, banks will raise the interest rates on loans to households and firms who wish to consume or invest. At higher interest rates the quantity demanded of products and capital for which households and firms must borrow decreases, as borrowers find higher interest rates less attractive. The opposite results from a fall in the price level and the decline in

interest rates, which makes borrowing more attractive and thus increases the quantity of output demanded.

- **The net export effect:** As the price level in a particular country falls, *ceteris paribus*, goods and services produced in that country become more attractive to foreign consumers. Likewise, domestic consumers find imports less attractive as they now appear relatively more expensive, so the net expenditures on the nation's goods rise as price level falls. The opposite results from an increase in the price level, which makes domestic output less attractive to foreigners and foreign products more attractive to domestic consumers.

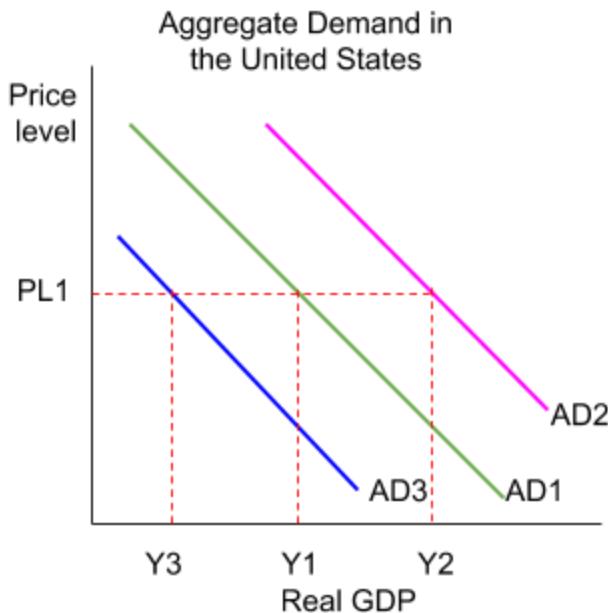
The components of aggregate demand

- Describe consumption, investment, government spending and net exports as the components of aggregate demand.

A change in the average price level of a nation's output will cause a movement along the AD curve, but a change in other variables will cause the AD curve to shift in or out. The components of aggregate demand are the four types of national expenditures. These are:

- Household **Consumption**
- Capital **Investment** by firms
- **Government spending**
- **Net Exports** (exports - imports)

A change in any of the four expenditures above will cause a shift in the AD curve. In the graph below, three levels of AD are shown. Notice that at each level of aggregate demand, a different quantity of real GDP is demanded in the country, even as the price level remains at PL1. (Note: The abbreviation "Y" is for "national income", another way of describing a country's GDP)



Assume a nation's economy begins at AD1:

- An increase in household consumption, business investment, government spending, or net exports will cause AD to increase to AD2. More output (Y_2) is demanded at the original price level of PL1.
- A decrease in any of the four expenditures will cause AD to decrease to AD3. Less output (Y_3) is demanded at the original price level of PL1.

The determinants of AD or causes of shifts in the AD curve

- Explain how the AD curve can be shifted by changes in consumption due to factors including changes in consumer confidence, interest rates, wealth, personal income taxes (and hence disposable income) and level of household indebtedness.
- Explain how the AD curve can be shifted by changes in investment due to factors including interest rates, business confidence, technology, business taxes and the level of corporate indebtedness.
- Explain how the AD curve can be shifted by changes in government spending due to factors including political and economic priorities.
- Explain how the AD curve can be shifted by changes in net exports due to factors including the income of trading partners, exchange rates and changes in the level of protectionism.

There are different factors that determine the level of aggregate expenditures in an economy.

Determinants of consumption (C)

Consumption refers to all the spending done by households on goods and services. The level of consumption in a nation depends on several factors.

The determinants of Consumption (C):	
Disposable Income	As disposable income rises, C increases. If disposable income falls, C will fall.
Wealth	When value of existing wealth (real assets and financial assets) increases, households tend to spend more on goods and services. When wealth decreases, consumption decreases.
Expectations	If households expect prices or their incomes to rise in the future, then today C increase, shifting AD out. If they expect lower prices or incomes, then C will likely decrease, as households choose to save more for the hard times ahead.
Real Interest Rates	Lower real interest rates lead to more C, as savings becomes less appealing and borrowing to buy durable goods can be done more cheaply.
Household Debt	When consumers increase their debt level, they can consumer more in the short-run. But if household debt is too high, C will eventually decrease

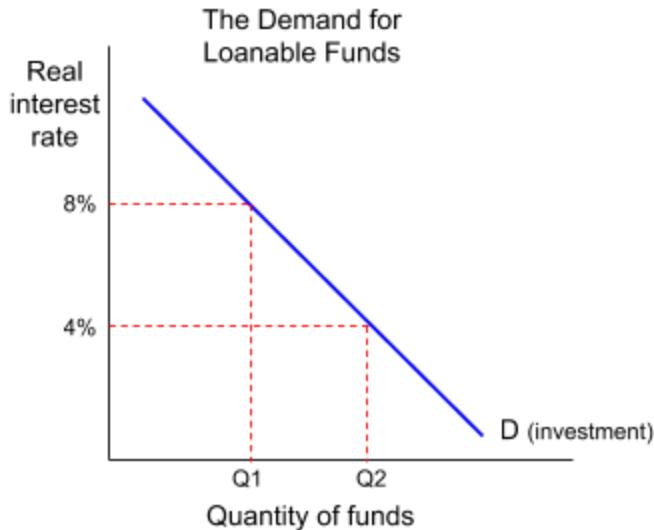
Taxation	Higher taxes decrease disposable income and causes C to fall. A decrease in taxes increases consumption. Taxes are set by government as part of fiscal policy.
-----------------	--

The determinants of investment (I)

Investment is defined as spending by firms on capital equipment or technology and by households on new homes. Changes in any of the following factors will cause a change in the level of investment in a country.

The determinants of Investment (I):	
The Real Interest Rate	Interest is the cost of borrowing money. Firms will borrow more to invest in new capital when the interest rate is low, and invest less when interest rates are high.
Business Confidence	If firms are confident about the level of future demand for their products, they are more likely to invest now. If confidence is low, firms will withhold from making new investments
Technology	New technology tends to spur new business investment, as firms rush to keep their manufacturing techniques as modern and efficient as possible and to produce the latest goods and services that consumers are demanding.
Business taxes	When firms can keep a larger share of their revenues (i.e. when taxes are lower) they may invest more. Higher business taxes discourage new investments.
The degree of excess capacity	If a firm's factories have excess capacity (meaning they are currently producing below the level they are capable of) firms will be less likely to invest since output can be increased without acquiring new capital.
Expectations:	If firms expect prices of their goods to be higher in the future, they are more likely to invest now. If lower prices are expected, firms have less incentive to invest now.

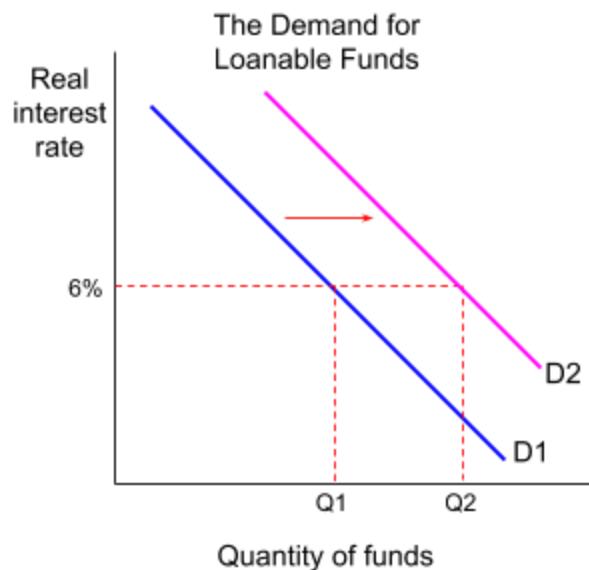
The **investment demand curve** shows the relationship between the real interest rate in a country and the quantity of funds demanded for investment by firms and households.



There is an inverse relationship between the real interest rate and the quantity of funds demanded for investment in a country.

- With a real interest rate of 8%, few firms will want to invest in new capital as there are very few investments with an expected rate of return greater than 8%
- When real interest rates are 4%, the quantity of funds demanded for investment is much higher, since there are more projects with an expected rate of return greater than 4%
- Firms will only invest if they expect the returns on the investment to be greater than the real interest rate (marginal benefit must be greater than the marginal cost)

If any of the determinants of investment change, there is a shift in the investment demand curve. For example, assume there is a new technology that spurs firms to invest in new capital equipment in order to meet the changing demands of consumers.



New technology has spurred firms to invest in new capital even as the interest rates has remained constant at 6%. As a result of the increase in investment spending, aggregate demand in the economy increases.

A decrease in one of the determinants of investment causes investment demand to shift left and results in less investment at every interest rate. For example, if the degree of excess capacity in the economy increases, firms will have less incentive to acquire new capital, causing the level of investment to decrease at every interest rate and a fall in aggregate demand.

The determinants of net exports (X_n)

Net exports is a component of aggregate demand that includes the revenue a country earns from selling its goods to the rest of the world (exports) minus the money the country spends on goods produced in other countries (imports).

$$\text{Net exports} = \text{Exports (X)} - \text{Imports (M)}$$

$$X_n = X - M$$

A change in net exports will cause a shift in aggregate demand. The following are the factors that can cause net exports to increase or decrease.

The Determinants of Net Exports (X_n):	
Foreign and Domestic Incomes	If the incomes of households in other nations rise, then demand for a country's exports should increase and net exports should rise. On the other hand, if domestic incomes rise, demand for imports will increase and net exports will fall.
The Exchange Rate	The exchange rate is the value of a country's currency relative to other currencies. As the exchange rate increases, a country's goods become more expensive and therefore less attractive to foreign consumers, while imports become cheaper, causing net exports to fall.
Protectionism	Protectionism refers to policies put in place by government intended to reduce amount of trade between one nation and others. Reducing protectionism will increase demand for imports in a country, and may cause net exports to fall. On the other hand, it may also increase demand for exports, causing net exports to increase.
Tastes and preferences	If a country's goods become more appealing to foreign consumers, demand for them will rise and net exports will increase.

The determinants of government spending (G)

Government spending is the final component of aggregate demand. The level of government spending in a nation is determined by the government's fiscal policy.

Fiscal policy refers to changes in the levels of taxation and government spending meant to increase or decrease the level of aggregate demand in a nation to promote macroeconomic objectives such as full employment, price level stability and economic growth. There are several key concepts relating to fiscal policy to consider:

- **The government's budget:** Changes in the level of government spending are reflected in the government's budget, towards which tax revenues go in order to finance government expenditures.
- **Contractionary fiscal policy:** If the government reduces its spending and / or increases the level of taxation, aggregate demand will decrease and shift to the left
- **Expansionary fiscal policy:** If the government increases its expenditures and / or decreases the level of taxation, aggregate demand will increase and shift to the right.
- **Budget deficit:** If, in a particular year, a government runs a deficit, meaning its expenditures exceed tax revenues, then government spending will contribute to AD that year and cause it to increase and shift out.
- **Budget surplus:** If, in a particular year, a government runs a surplus, meaning it spends less than it collected in taxes, then fiscal policy will subtract from aggregate demand, shifting it to the left.

The meaning of aggregate supply

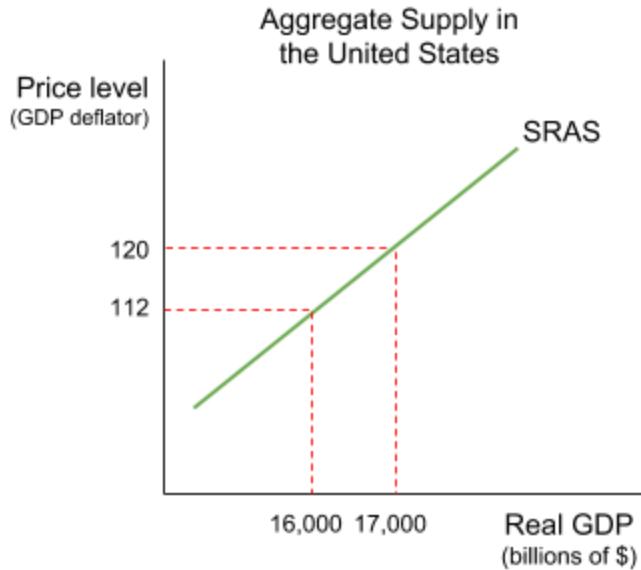
- Describe the term aggregate supply.
- Explain, using a diagram, why the short-run aggregate supply curve (SRAS curve) is upward sloping.
- Explain, using a diagram, how the AS curve in the short run (SRAS) can shift due to factors including changes in resource prices, changes in business taxes and subsidies and supply shocks.

Aggregate supply is the total quantity of output of goods and services produced by the firms in a nation at a range of price levels in a particular period of time.

Similar to supply in microeconomics, firms respond to higher prices by increasing their output in the short-run and to lower price levels by decreasing their output in the short-run. The difference between supply and aggregate supply, of course, is that AS shows the total output against the average price level while supply shows the quantity of a particular good against that good's price.

Short-run aggregate supply (SRAS) illustrates the relationship between a country's aggregate price level of and the quantity of total output produced in the period of time over which workers' wages and other input prices are fixed (or **sticky**). In other words, it shows how firms will respond to price changes when the prices of their raw materials and labor cannot be changed due to contracts and other legal constraints.

Because of sticky wages and input prices, the SRAS is upward sloping.



From the SRAS curve above we can see the following:

- The aggregate price level is measured using a price index (in this case, the GDP deflator)
- At higher price levels (120) the amount of output supplied by the nation's firms (\$17,000 billion) is greater than at lower price levels (112 on the index and \$16,000 billion worth of output)
- There is a direct relationship between the aggregate price level and the amount of output produced in the short run.

Sticky wages and input prices explained

The shape of the SRAS curve reflects the fact that the prices of inputs (raw materials, rent for land, interest on capital, electricity prices, transport and fuel costs) and worker wages are sticky, or inflexible in the short run. Here's why: if all these prices (which make up the costs of production for firms) were perfectly flexible, they would quickly adjust to whatever the level of demand in the economy is and firms would raise and lower them as demand for their output changed. In such a world, there would be no reason for firms ever to change their level of output, and the economy would always produce at an efficient level where all resources are full employed.

However, this is not the world we live in, and in the real world input prices and wages are sticky. Due to labor contracts, legal barriers, and other factors, firms cannot quickly slash their prices to adjust their costs in the face of falling demand; nor do all their costs increase immediately when there is increased demand for their output. As a result, firms will reduce output in order to reduce costs when demand for their stuff falls, and increase output when demand for their stuff increases.

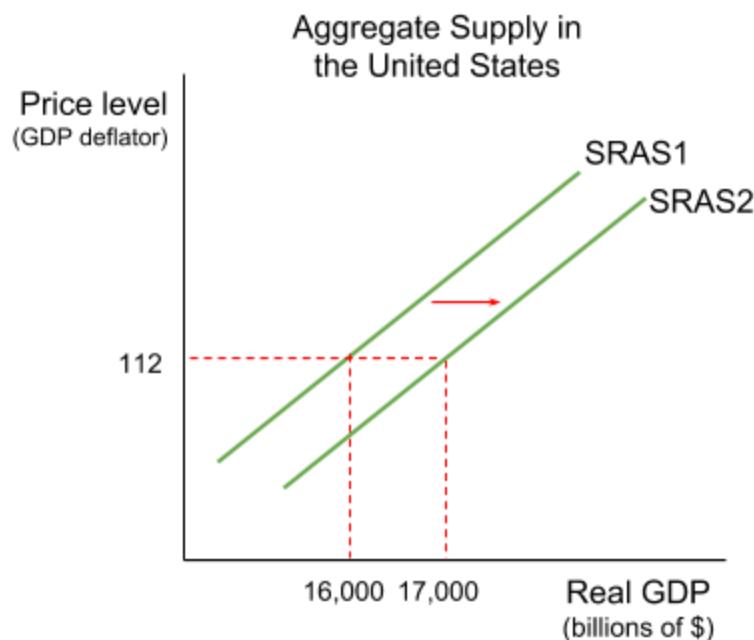
Factors that shift short-run aggregate supply

Shifts in SRAS result from anything that causes an increase or a decrease in the costs of production for a nation's firms. When any of the following change, aggregate supply will

either decrease and shift inwards (or up, graphically) or increase and shift outwards (or down, graphically).

- **Wage rates:** The cost of labor. Higher wages cause SRAS to decrease, lower wages cause SRAS to increase
- **Resource costs:** Rents for land, interest on capital; as these rise and fall, SRAS will shift in or out
- **Energy and transportation costs:** Higher oil or energy prices will cause SRAS to decrease. If costs fall, SRAS increases
- **Government regulation:** Regulations impose costs on firms that can cause SRAS to decrease
- **Business taxes/subsidies:** Taxes are a monetary cost imposed on firms by the government, and higher taxes will cause SRAS to decrease. Subsidies reduce firms' costs and cause SRAS to increase.
- **Exchange rates:** If a country's producers use lots of imported raw materials, then a weaker currency will cause these to become more expensive, reducing SRAS. A stronger currency can make raw materials cheaper and increase SRAS.

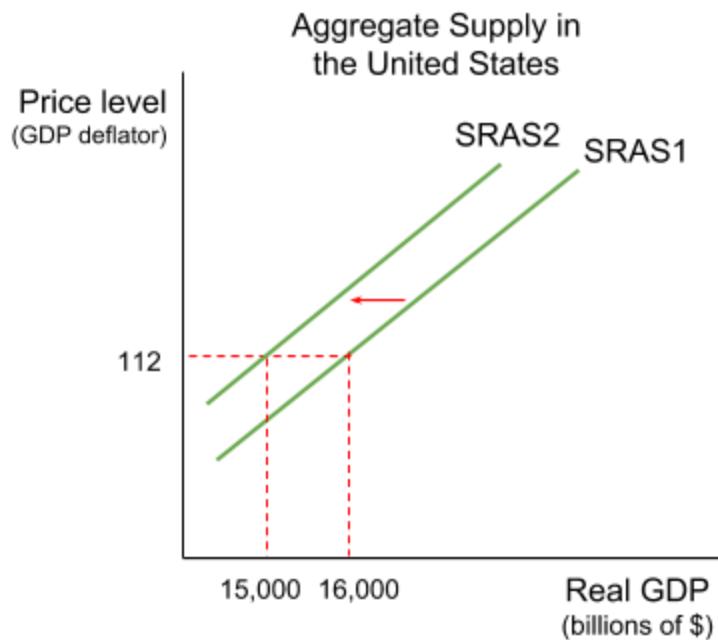
A shift in SRAS results in a larger or smaller amount of output being produced at every price level. For example, examine the graph below.



Assume there has been a decrease in energy prices in the United States.

- As a result of lower energy prices, firms' costs have decreased and SRAS has shifted out.
- At the original price level (where the GDP deflator price index is 112), there is now more output possible than before.
- Firms are willing to produce \$17,000 billion worth of output at a price index of 112 compared to only \$16,000 billion before energy prices fell.

Next, assume something changes to cause SRAS to decrease, as in the graph below.



Assume there has been an increase in the business tax rate.

- Due to higher taxes, firms must pay the government a larger amount for every unit of output they sell.
- Firms are willing to produce less output at the original price level (where the GDP deflator price index is 112).
- The amount of output produced has decreased due to higher business taxes. Firms are only willing to produce \$15,000 billion worth of goods and services compared to \$16,000 billion before taxes were increased.

The tradeoff between inflation and unemployment in the SRAS

When the aggregate price level increases, output increases in the short run. At higher levels of output, firms employ more workers; therefore, unemployment decreases as the price level increases.

When the aggregate price level decreases, output decreases in the short run. At lower levels of output, firms employ fewer workers; therefore, unemployment increases as the price level decreases.

Due to the stickiness of wages and other input costs, there is an inverse relationship between inflation and unemployment in the short run. As the price level increases, unemployment decreases and as the price level decreases unemployment increases.

The relationship between unemployment and inflation will be revisited in a later unit when we learn about the Phillips Curve.

Alternative views of aggregate supply

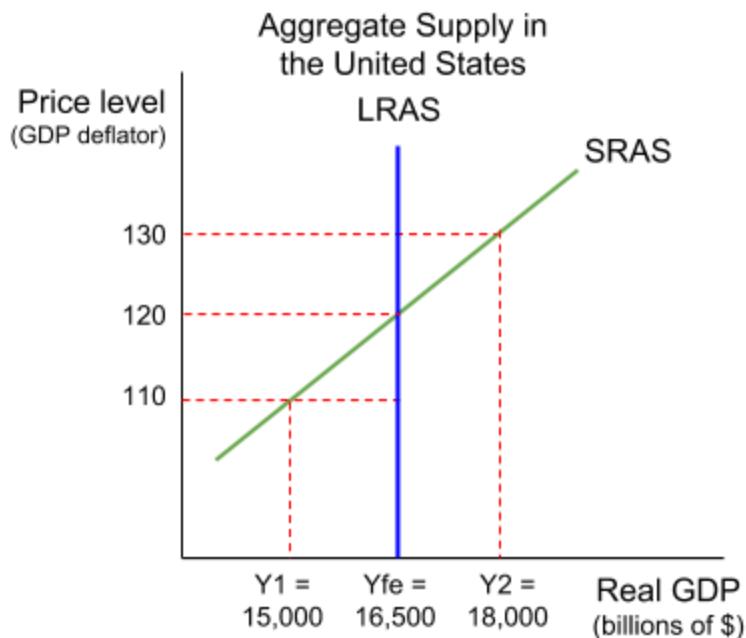
- Explain, using a diagram, that the monetarist/new classical model of the long-run aggregate supply curve (LRAS) is vertical at the level of potential output (full employment output) because aggregate supply in the long run is independent of the price level.
- Explain, using a diagram, that the Keynesian model of the aggregate supply curve has three sections because of “wage/price” downward inflexibility and different levels of spare capacity in the economy

Introduction to long-run aggregate supply (LRAS)

We have already defined the short run in macroeconomics as the “sticky-wage period”: the period of time over which wages and other input costs cannot be changed, due to contracts and other legal constraints.

The long run in macroeconomics is the period of time over which all input costs (wages, raw materials, rents, etc.) are fully flexible and will adjust to the price level in the economy. In the long run, the level of output produced by a nation’s firms is independent of the price level; output will always be at the full employment, potential level of output in the long run, since any change in the price level will be met by a change in input costs.

The **long-run aggregate supply curve (LRAS)** is therefore vertical at the potential, full employment level of output. The graph below shows the SRAS and the LRAS for the United States.



The graph above shows three price levels (GDP deflators of 110, 120, and 130), and three different levels of short-run output (Y_1 , Y_{fe} , and Y_2). Notice the following:

- At a price level of 110, short-run output in the United States is at a level below full employment (denoted by Y_{fe}), since at lower price levels, firms must reduce output in the short run due to the sticky nature of wages and other input prices.
- At a price level of 130, short-run output is at a level beyond full employment, since at higher price levels firms will increase their output due to sticky wages and other input prices.
- At a price level of 120, short-run output is at the full employment.
- In the long run, output is always at the full employment level of \$16,500 billion.

LRAS is vertical at the full employment level of output because wages and other input prices are fully flexible in the long run. For example, assume there is a fall in the price level from 120 to 110.

In the short run:

- Firms have labor contracts, rental agreements, and contracts with suppliers that have locked in the wages, rents, and other input costs they face.
- When prices start to fall, firms must cut their costs to remain in business. But since input prices are fixed, the only way to cut costs is to reduce output.
- In the short run, falling prices lead firms to reduce output and employment, causing a **negative output gap** as output falls from \$16,500 billion to \$15,000 billion.

In the long run:

- High unemployment leads workers to accept lower wages from employers. At the same time, rental contracts and prices from input suppliers can be renegotiated at lower prices.
- Input costs fall, and firms begin increasing output and lowering their prices further.
- In the long run, falling input prices result in output returning to full employment, as firms pass their lower costs on to consumers and increase production and lower prices.

Because wages and other input costs will adjust to the price level in the long run, output will return to full employment.

The reverse happens when the price level increases to 130.

In the short run:

- Firms' input costs are fixed, so they increase their output to chase the profits promised by higher prices.
- Higher prices lead to increased output and employment, and a **positive output gap** is created.

In the long-run:

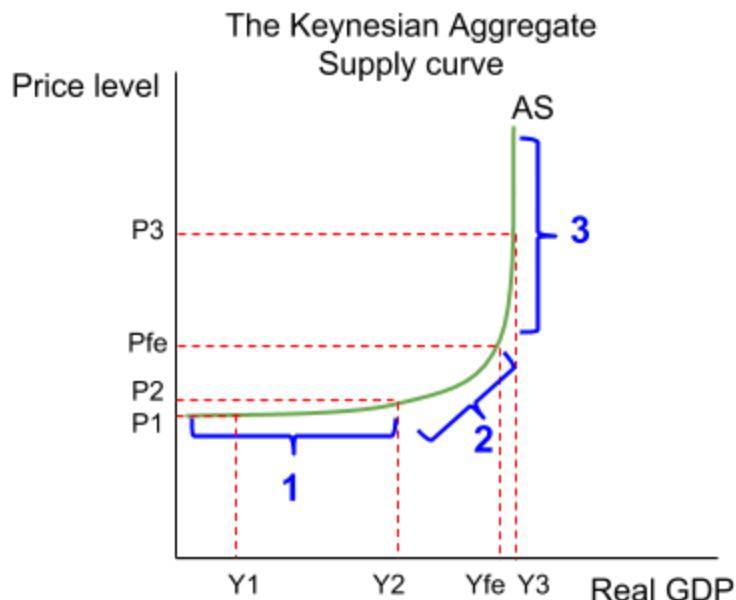
- Low unemployment drives wages up, and landlords, banks, and raw material suppliers negotiate their rents, interest rates, and other input prices, raising firms' input costs.

- As resource costs rise in the long run, firms reduce their output once more until GDP returns to the full employment level.

In the long run, there is no tradeoff between inflation and unemployment. Because wages will always adjust to the price level, output will always return to the full employment level, which corresponds with the **natural rate of unemployment**.

Another view of aggregate supply holds that there is no difference between short-run and long-run aggregate supply, rather there is a single AS curve with three different sections depicting downward wage and price-inflexibility and different levels of spare capacity in the economy.

The **Keynesian model of aggregate supply**, proposed by Depression-era economist John Maynard Keynes, is horizontal (perfectly elastic) below full employment but vertical (perfectly inelastic) beyond full employment.



In the Keynesian model, the three sections (1, 2, and 3) can be understood as follows:

- Section 1:** There is a large degree of spare capacity in the economy. Unemployment is high and there is ample unused land and capital. Increases in output (from Y_1 to Y_2) are accompanied by minimal inflation (P_1 to P_2).
- Section 2:** The economy is approaching full capacity. Unemployment is low and land and capital are becoming scarcer. Further increases in output (from Y_2 to Y_{fe}) will cause prices to rise as there is increased demand relative to the supply of resources, causing inflation to increase (from P_2 to P_{fe}).
- Section 3:** The economy is beyond its full employment level of output. Unemployment is extremely low and there is little or no spare capacity in the economy. All factories and productive land are being used. Further increases in

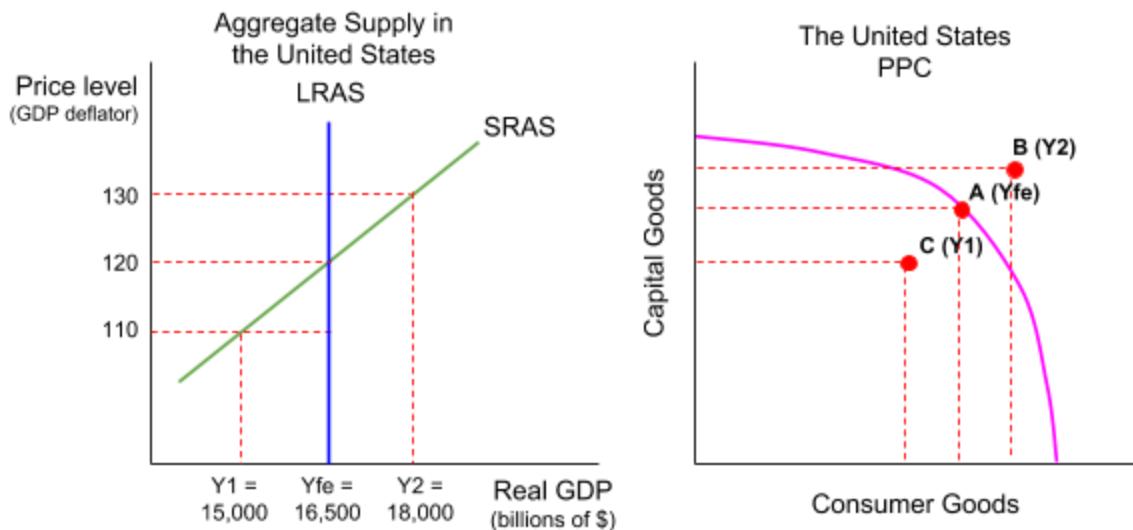
output (Y_{fe} to Y_3) are met only with runaway inflation (P_f to P_3).

For the bulk of our analysis in the coming units we will examine national output and price determination using the SRAS and LRAS models. However, most of the analysis and theory can be examined and applied using the Keynesian model of aggregate supply as well.

Shifting the aggregate supply curve over the long term

- Explain, using the two models above, how factors leading to changes in the quantity and/or quality of factors of production (including improvements in efficiency, new technology, reductions in unemployment, and institutional changes) can shift the aggregate supply curve over the long term.

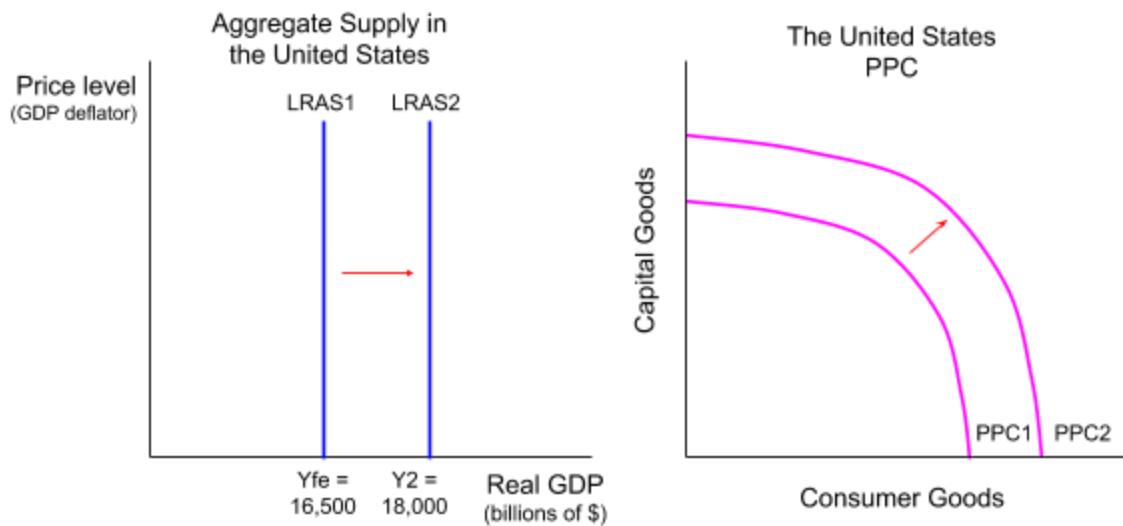
A nation's LRAS curve corresponds with its production possibilities curve (PPC). Both show the potential level of output a country can sustain in the long run, assuming resources are fully employed in the economy.



The PPC on the right corresponds with the LRAS on the left.

- At Y_1 the economy has a negative output gap. It is producing below full employment and inside its PPC.
- At Y_2 the economy has a positive output gap. It is producing beyond full employment and outside its PPC.
- In the long run, the economy is producing at full employment output and on its PPC. Output always returns to full employment (and to the PPC) in the long run because wages and other input costs adjust to the price level until resources are fully employed.

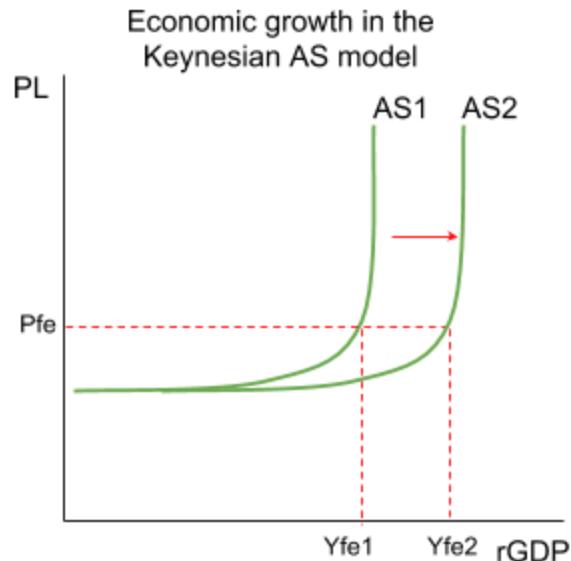
A change in the factors of production of a nation will shift its LRAS, causing either economic growth or contraction in the nation. Shifts in the LRAS correspond with shifts in the PPC.



Factors that can cause a shift in a nation's LRAS include:

- Changes in population (more workers lead to greater potential output, fewer workers lead to less potential output)
- Changes in technology (new technologies increase potential output)
- Increased trade (more imported resources increase potential output)
- Improvements in human capital (better educated or skilled workers increase potential output)

In the Keynesian AS model, a change in any of the factors above will likewise cause an outward shift in aggregate supply and an increase in a country's actual and potential output.

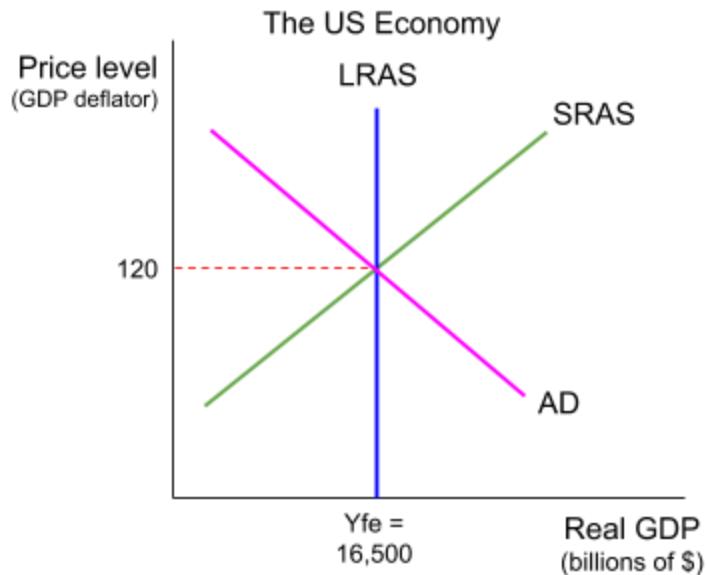


Short-run equilibrium in the AD/AS model

- Explain, using a diagram, the determination of short-run equilibrium, using the SRAS curve.
- Examine, using diagrams, the impacts of changes in short-run equilibrium.
- Explain, using the Keynesian AD/AS diagram, that the economy may be in equilibrium at any level of real output where AD intersects AS.
- Explain, using a diagram, that if the economy is in equilibrium at a level of real output below the full employment level of output, then there is a deflationary (recessionary) gap.
- Discuss why, in contrast to the monetarist/new classical model, the economy can remain stuck in a deflationary (recessionary) gap in the Keynesian model.
- Explain, using a diagram, that if AD increases in the vertical section of the AS curve, then there is an inflationary gap.
- Discuss why, in contrast to the monetarist/new classical model, increases in aggregate demand in the Keynesian AD/AS model need not be inflationary, unless the economy is operating close to, or at, the level of full employment

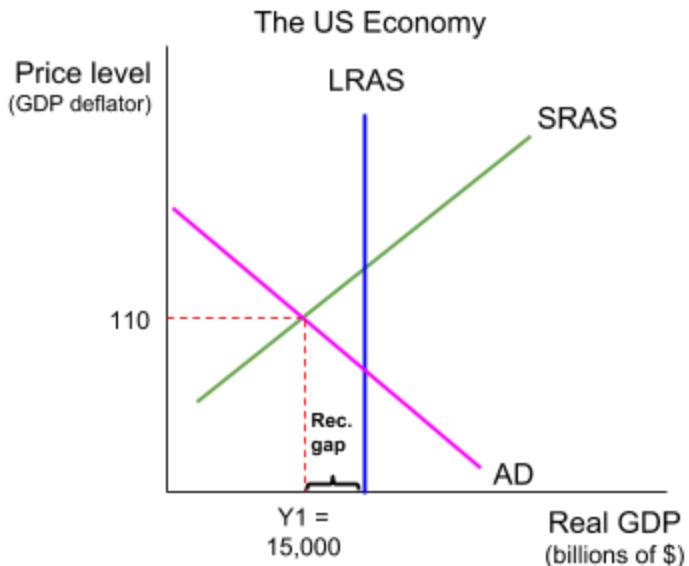
A country's **short-run equilibrium level of national output** refers to its output at any particular period of time, and occurs at the price level and real GDP at which AD and SRAS intersect.

Short-run equilibrium output can be at the full-employment level of output or it can beyond full-employment or below full-employment. In the graph below, a country is producing a short-run equilibrium level of output that is equal to the full employment level.



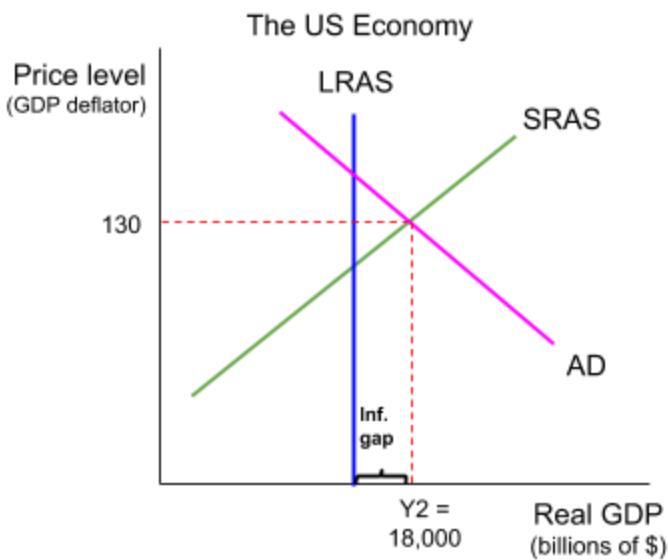
The country is producing its full employment level of output. The unemployment rate in this economy is equal to the natural rate of unemployment. This country producing at its **long-run equilibrium level of output**, which is where $AD=SRAS=LRAS$.

Short-run equilibrium can also occur below full employment, as in the graph below.



The country is producing below full employment and has a **recessionary gap** (a negative output gap). The unemployment rate is higher than the natural rate of unemployment. There is **cyclical unemployment** in this country, as it is experiencing a recession.

Finally, a country can produce a short-run equilibrium level of output that is beyond full employment.



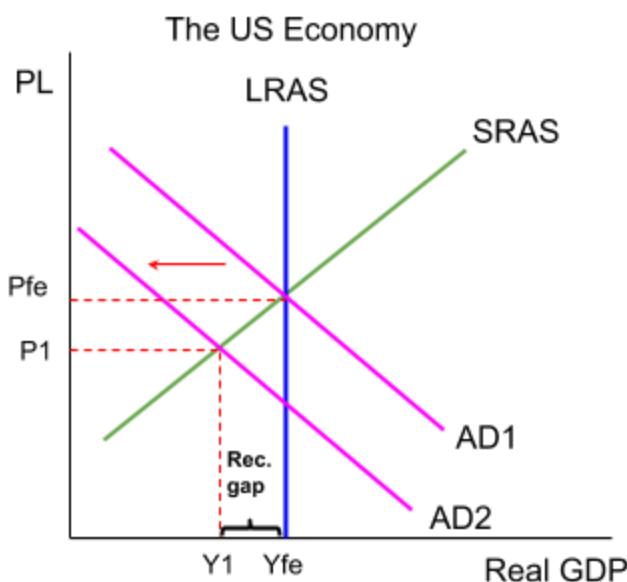
The country is producing beyond full employment and has an **inflationary gap** (a positive output gap). The unemployment rate is lower than the natural rate of unemployment (in other words, there is negative cyclical unemployment). This economy is overheating;

resources are stretched to their limit as the economy approaches a peak in its business cycle.

Aggregate Demand shocks

An **aggregate demand shock** occurs when one of the components of aggregate demand increases or decreases. For example, assume there is an increase in income taxes, which causes household consumption to decrease across the economy.

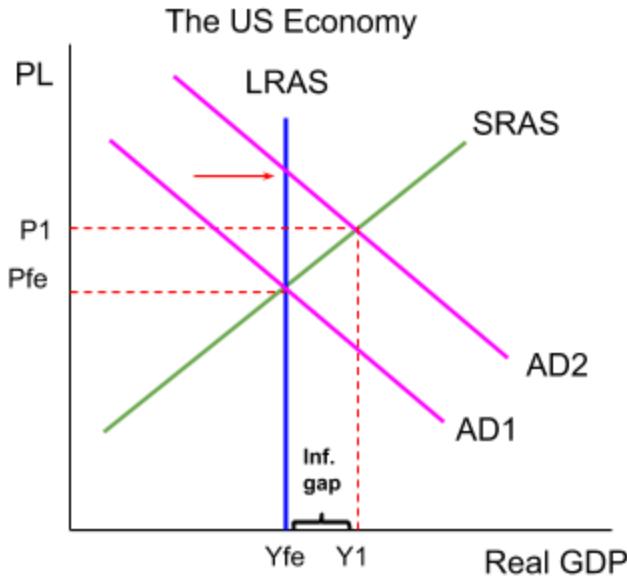
The graph below shows the effect of a **negative demand shock** on short-run equilibrium in the AD-AS model.



The decrease in consumption has caused a negative demand shock in the United States:

- The price level has decreased from P_{fe} to P_1 (this may be disinflation or deflation, depending on the magnitude of the decrease in AD)
- Firms, which are unable to lower the wages and other input prices they pay, must reduce output, which falls from Y_{fe} to Y_1 , causing a recessionary gap.
- Cyclical unemployment is created, causing the unemployment rate to increase to a level greater than the NRU

Assume next that there is a decrease in interest rates, and interest sensitive investment and consumption both increase, causing a **positive demand shock**. The effect on short-run equilibrium can be seen below:

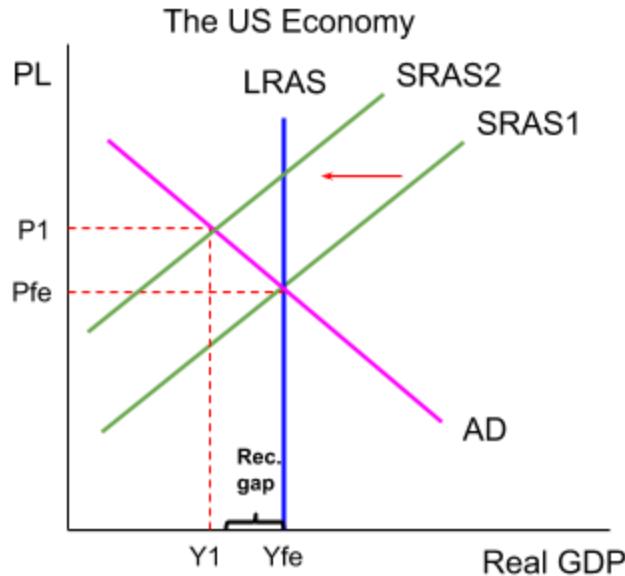


The increase in investment and consumption has caused a positive demand shock in the United States:

- The price level has increased from P_{fe} to P_1 (meaning the inflation rate has increased)
- Firms, which can still pay workers the fixed wage rate they had previously negotiated and can still get other inputs at relatively cheap prices, increase their output to chase the higher profits made possible due to higher prices. The economy experiences an inflationary gap.
- Unemployment decrease to a rate below the NRU, as structurally and frictionally unemployed workers are hired to meet the rising demand from firms for new capital and from households for new products.

Aggregate Supply shocks

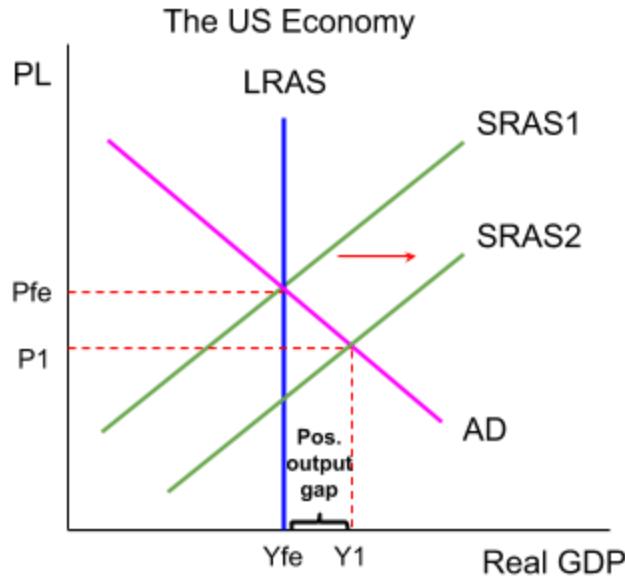
An **aggregate supply shock** occurs whenever one of the determinants of SRAS changes, causing a shift in the SRAS curve. For example, assume there is a massive increase in government regulation of businesses, resulting in increased costs among the nation's firms as they begin to comply to new, stricter, government rules.



Increased regulation has caused a **negative supply shock** in the United States:

- The higher costs to comply with stricter government rules has caused the SRAS to decrease and the price level to increase. Inflation resulting from higher costs among firms is called **cost-push inflation**.
- Firms must reduce their output in the short run as other wages and other costs are fixed and increased government regulation increases the cost of adhering to government rules. The economy experiences a recessionary gap as output falls to a level below full employment.
- Workers are laid off as firms' costs rise and output falls. The level of unemployment increases to a rate higher than the NRU.

Something that increases SRAS will cause a **positive supply shock**. Assume, for example, there is an unexpected drop in energy prices due to a new, cheaper technology for producing electricity. The resulting positive supply shock will have the following effect on short-run equilibrium in the economy.



Cheaper electricity has reduced the cost of producing most goods and services, causing SRAS to increase:

- The price level decreases as firms pass their lower energy prices onto consumers (this could mean disinflation or deflation, depending on the magnitude of the shift in SRAS).
- Firms increase their output as the price of electricity (a major cost for many firms) decreases and producing more output becomes increasingly profitable.
- Output increases beyond the full employment level and employment increases, causing unemployment to fall below the natural rate of unemployment.
- If the lower energy prices are permanent, the LRAS curve will shift out and Y_1 will become the country's new full employment level of output. However, if energy price increase again, the SRAS will shift left and output will return to its original full employment level (Y_{fe}).

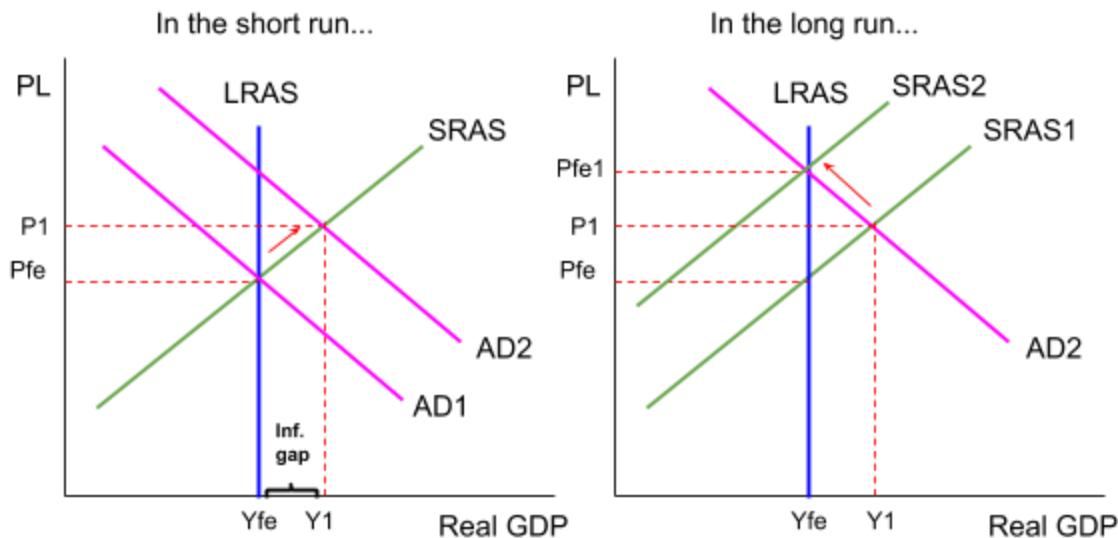
Long-run equilibrium in the AD/AS model

- Explain, using a diagram, the determination of long-run equilibrium, indicating that long-run equilibrium occurs at the full employment level of output.
- Explain why, in the monetarist/new classical approach, while there may be short-term fluctuations in output, the economy will always return to the full employment level of output in the long run.
- Examine, using diagrams, the impacts of changes in the long-run equilibrium.

From short run to long run in the AD-AS model

In the short run wages and other costs of production are fixed, which explains why firms respond to changes in AD and the price level by reducing output (when AD falls) and increasing output (when AD increases). However, in the long run, wages and other input costs will fully adjust to the price level and to AD, restoring full employment.

Consider the US economy following the decrease in interest rates that caused a positive demand shock and demand-pull inflation. The graphs below show the series of events that will restore full employment in the economy in the long run, once wages and other input prices have adjusted to the higher demand.



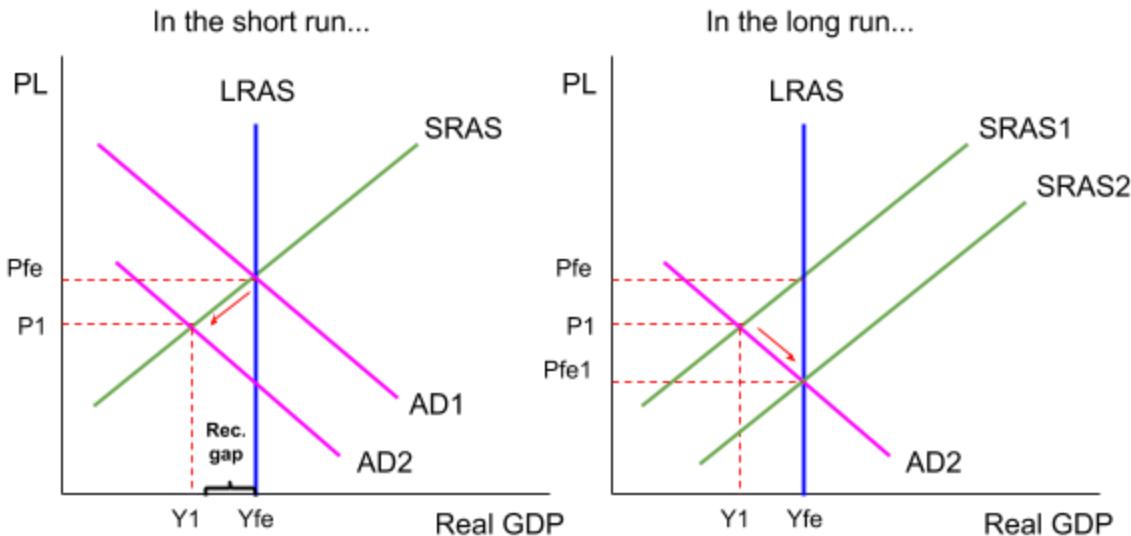
Following the positive demand shock, the economy experiences a short-run increase in output and inflation, but in the long run:

- Wages increase due to unemployment that is below the NRU; workers have more bargaining power and will negotiate pay increases from current and prospective employers.
- Raw material prices, rents, interest rates, and other input prices will increase as these resources become increasingly scarce due to an economy that is producing beyond its full employment level.
- Firms will face higher wages and other input costs, forcing them to reduce output and raise prices even further.
- SRAS shifts in due to higher resource costs, causing output to return to full employment (Y_{fe}).
- Inflation is now higher than following the initial increase in demand, as the economy has now experienced both demand-pull inflation and cost-push inflation.
- Unemployment returns to its natural rate.

As can be seen, in the long run it is not possible for an economy to produce beyond its full employment level of output, because wages and other input costs will eventually adjust to the higher price level and level of AD, causing output to return to its full employment level.

Next let's consider how an economy will adjust in the long run following a decrease in aggregate demand. Assume that America's major trading partners experience recessions and as a result there is a fall in demand for US goods overseas. The resulting decrease in net

exports will cause a negative demand shock. The graphs below show the series of events that will restore full employment in the economy in the long run, once wages and other input prices have adjusted to the lower level of AD.



Following the negative demand shock, the economy experiences a short-run decrease in output and inflation (maybe even deflation), but in the long run:

- Wages decrease due to unemployment that is greater than the NRU; firms have more bargaining power and will negotiate pay cuts for current and prospective employees.
- Raw material prices, rents, interest rates and other input prices will decrease as these resources become less and less scarce due to an economy that is producing below its full employment level.
- Firms will face lower wages and other input costs, allowing them to increase output and reduce prices even further.
- SRAS shifts out due to lower resource costs, causing output to return to full employment (Y_{fe}).
- Inflation is now lower than following the initial increase in demand. This may mean the economy experiences lower inflation year on year than it did before the recession or perhaps prices overall are lower than they were before the recession.
- Unemployment returns to its natural rate.

In the long run, in the absence of government policy actions, flexible wages and prices will adjust to restore full employment and unemployment will revert to its natural rate after a shock to aggregate demand.

Economic growth occurs when there is an increase in a country's actual and potential output of goods and services over time. From our analysis of changes in the AD-AS model, we can conclude the following:

- Changes in AD alone will not cause economic growth. A decrease in AD causes a recession in the short-run, and an increase in AD causes demand-pull inflation and

- an overheating economy. In both cases, output returns to full employment once wages and other input prices have fully adjusted.
- Changes in SRAS can cause a short-run decrease in output (if resource costs increase) or increase in output (if resource costs decrease).

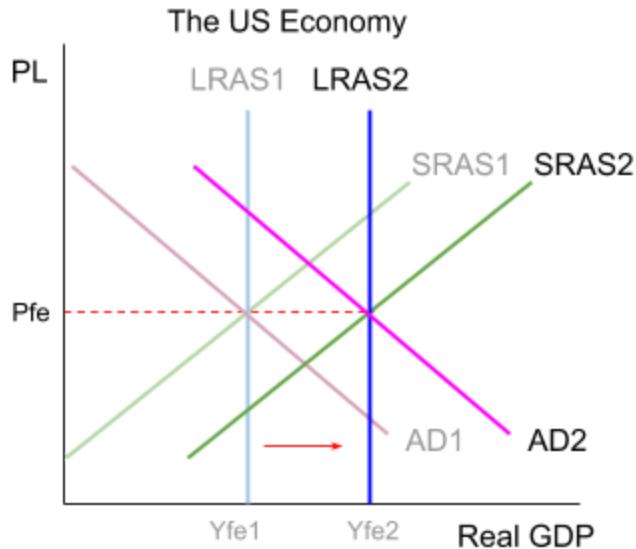
In other words, shifts in AD and SRAS alone will not cause an increase in an economy's potential output, only in its actual output (its short-run equilibrium real GDP). In order for potential output to increase, there must be an outward shift in the long-run aggregate supply (LRAS) curve.

LRAS will increase if there is an increase in any of the following:

- The quantity of factors of production:
 - More workers
 - More capital
 - More land resources
- The quality of the factors of production:
 - Better human capital:** A more educated or higher skilled workforce increases a country's potential output.
 - Greater productivity:** More output per worker (usually the result of new or better technology) increases potential output.
 - Technological advancement:** Improvements in technology lead to more potential output in the sectors in which they are experienced. For example, better farming equipment increases potential output in the agricultural sector; faster internet increases potential output in the knowledge/IT sectors, and so on.

Increases in a nation's potential output through advances in human capital or technology cause the LRAS curve to shift out. Typically, such changes also result in increased AD and SRAS as well (due to new investment and consumption or increased government spending on human or physical capital).

Assume that due to a combination of increased immigration and improvements in the education system, the human capital of the United States has increased both in quantity and quality. More, better skilled workers increase AD (more consumers) and both SRAS and LRAS, as in the graph below.



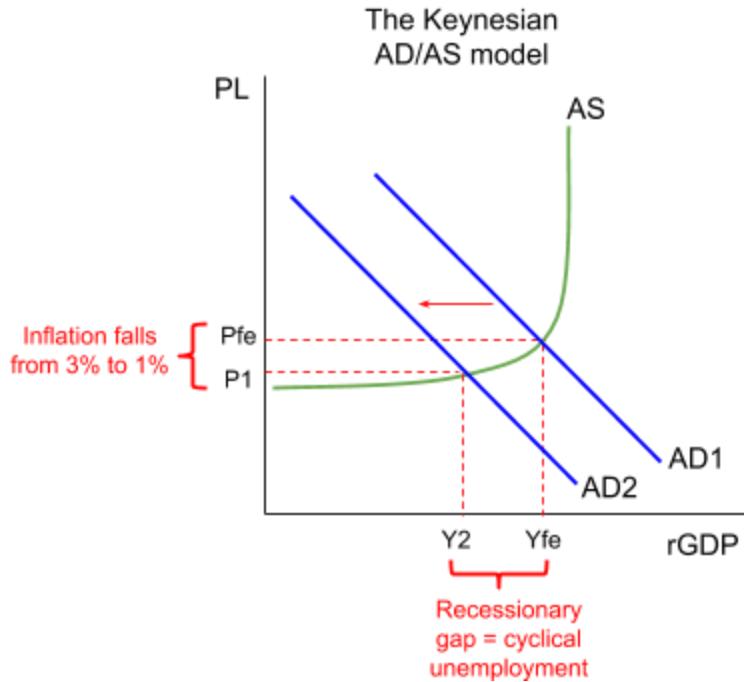
More and better skilled workers has lead to an increase in both the short-run equilibrium level of output and the full employment level of output. The country experiences economic growth as the LRAS curve shifts outward.

Short-run and long-run equilibrium in the Keynesian AS model

The SRAS/LRAS/AD model in our above analysis assumes that wages and other input prices will ultimately adjust to demand shocks: when AD is low, wages and prices fall, restoring full employment output as SRAS shifts out. When AD is high, wages and prices rise, restoring full employment as SRAS shifts in.

In contrast to this “flexible wage” view of the economy, the economy can remain stuck in a deflationary/recessionary gap in the Keynesian model.

The graph below shows an economy in which AD has fallen (perhaps due to reduced export demand from abroad) in the Keynesian model.

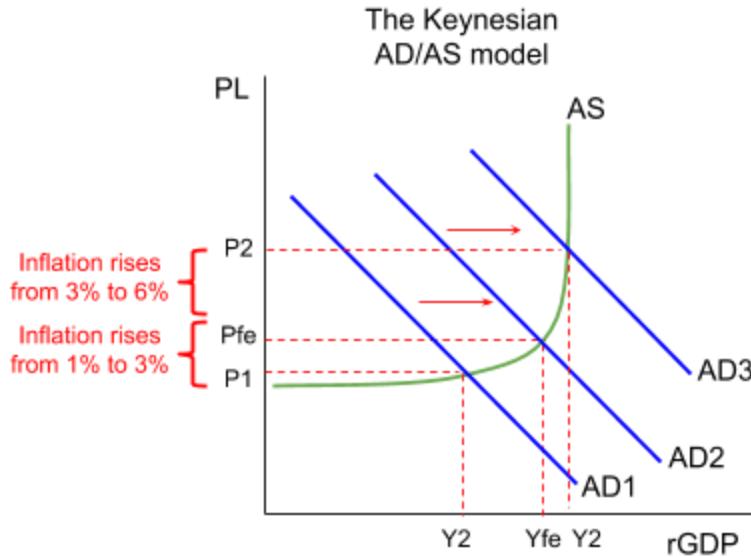


The Keynesian model assumes that wages and prices are sticky. Rather than experiencing deflation, there is disinflation (a lower inflation rate). Cyclical unemployment results as firms must cut costs and reduce output due to falling demand.

Since prices don't fall, wages will remain sticky and the economy can remain stuck in the recessionary gap caused by the fall in AD. In order to restore full employment, according to the Keynesian view, aggregate demand must increase.

Keynesians believe government should take an active role in managing the level of aggregate demand in the economy to maintain full employment. Recessionary gaps should be closed through active “demand-side policies”, including fiscal and monetary policies, which will be explored in a later chapter.

Increases in AD, according to the Keynesian model, need not be inflationary, unless the economy is operating close to, or at, the level of full employment. Assume AD increases from AD1 to AD2 to AD3 as in the model below.



The increase from AD1 to AD2 is accompanied by only a slight increase in inflation, as there is excess capacity in the economy, allowing firms to hire workers and employ more capital and land without driving costs up.

However, the increase from AD2 to AD3 drives inflation above the desired rate of 3% due to the lack of spare capacity in the economy. Workers, capital, and land are all scarcer at AD3 and the slight increase in output (Y_{fe} to Y_2) is accompanied by a large increase in inflation (3% to 6%).

According to the Keynesian model, sticky wages and prices mean that aggregate supply is highly elastic below full employment (large increases in output are accompanied by only small increases in inflation) and highly inelastic beyond full employment (small increases in output are accompanied by large increases in inflation).

The Keynesian spending multiplier (HL only)

- Explain, with reference to the concepts of leakages (withdrawals) and injections, the nature and importance of the Keynesian multiplier.
- Calculate the multiplier using either of the multiplier formulae.
- Use the multiplier to calculate the effect on GDP of a change in an injection in investment, government spending or exports.
- Draw a Keynesian AD/AS diagram to show the impact of the multiplier.

Introduction to the multipliers

A change in one of the four types of spending of a certain amount will always lead to a larger change in total expenditures and aggregate demand. For example, if government increases its infrastructure spending by \$1 billion, we would expect total spending in the economy (GDP) to increase by more than \$1 billion.

The multiplier effect refers to the fact that a particular change in spending or taxation in the economy will result in a larger change in overall aggregate demand.

Whenever households experience a change in their income, both the levels of injections and leakages in the economy will change. Households basically do four things with their incomes:

- Spend it on domestically produced goods: the proportion of any change in income that is spent on domestically produced goods is known as the **marginal propensity to consume (MPC)**, which measures the change in consumption (C) resulting from a particular change in income (Y).

$$\text{MP C} = \frac{\Delta C}{\Delta Y}$$

- Save it: the proportion of any change in income that is saved is known as the **marginal propensity to save (MPS)**, which measures the change in household savings (S) resulting from a particular change in income (Y).

$$\text{MP S} = \frac{\Delta S}{\Delta Y}$$

- Pay taxes with it: the proportion of any change in income that is taxed is known as the **marginal propensity to tax (MPT)**, which measures the change in taxes (T) resulting from a particular change in income (Y).

$$\text{MP T} = \frac{\Delta T}{\Delta Y}$$

- Buy imports with it: the proportion of any change in income that is spent on imported goods is known as the **marginal propensity to import (MPM)**, which measures the change in import spending (M) resulting from a particular change in income (Y).

$$\text{MP M} = \frac{\Delta M}{\Delta Y}$$

Whenever there is a change in households' incomes resulting from a change in expenditures or taxes in a country, the four components above will all be affected to a degree determined by the marginal propensities.

For example, assume that the typical household spends 80% of any new income they earn on domestic goods, buy imports with 10% of it, saves 5% and pays 5% in taxes. The marginal propensities are thus as follows:

- MPC = 0.8
- MPM = 0.1
- MPS = 0.05
- MPT = 0.05

Notice that the sum of the marginal propensities is always equal to 1:

$$MPC + MPM + MPS + MPT = 1$$

$$0.8 + 0.1 + 0.05 + 0.05 = 1$$

When there is an initial change in aggregate expenditures in a country (C, I, G, or Xn), the ultimate change in aggregate demand can be calculated by multiplying the initial change by the **Keynesian spending multiplier**.

$$\text{Keynesian spending multiplier (k)} = \frac{1}{1-MPC}$$

or

$$k = \frac{1}{MPS + MPT + MPM}$$

For example, using the marginal propensities above, assume there is an increase in income from net exports in the United States of **\$1 billion**. The \$1 billion increase in net exports will increase household income by \$1 billion, which will lead to...

- \$800 million increase in spending
- \$100 million increase in imports
- \$50 million increase in savings
- \$50 million increase in taxes

The initial injection of \$1 billion has caused a secondary injection of an additional \$800 million, as consumption has increased in addition to the initial increase in net exports. In this way, the initial change in spending is multiplied as further increases in spending result, as the \$800 million increase in consumption causes further increases in consumption of...

- $\$800m \times 0.8 = \$640m$
- $\$640m \times 0.8 = \$512m$
- $\$512m \times 0.8 = \$409.6m$
- $\$409.6m \times 0.8 = \$327.68m$
- ...and so on

The ultimate change in total spending (GDP) resulting from an initial increase in aggregate expenditures can be calculated by multiplying the initial change in spending by the Keynesian spending multiplier:

$$\Delta GDP = \text{initial } \Delta \text{ in spending} \times \text{the spending multiplier}$$

$$= \$1 \text{ billion} \times \frac{1}{1-0.8}$$

$$= \$1 \text{ billion} \times \frac{1}{0.2}$$

$$= \$1 \text{ billion} \times 5 = \textcolor{red}{\$5 \text{ billion}}$$

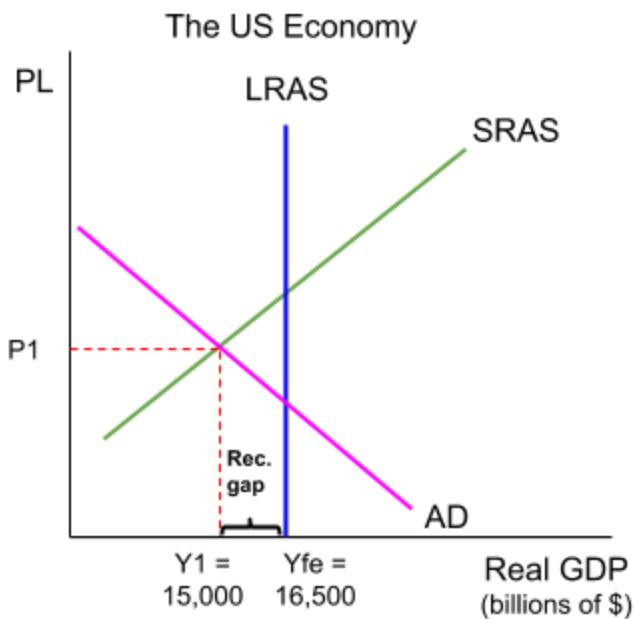
When households spend 80% of any change in their income, a \$1 billion increase in national income resulting from an increase in net exports will increase overall GDP by \$5 billion.

The larger the marginal propensity to consume, the larger the spending multiplier will be. The table below shows some possible values for the MPC and the corresponding sizes of the spending multipliers.

MPC	Keynesian spending multiplier	Effect on GDP of \$1b increase in AD
0.9	$\frac{1}{0.1} = 10$	\$1b × 10 = \$10b
0.6	$\frac{1}{0.4} = 2.5$	\$1b × 2.5 = \$2.5b
0.5	$\frac{1}{0.5} = 2$	\$1b × 2 = \$2b
0.4	$\frac{1}{0.6} = 1.67$	\$1b × 1.67 = \$1.67b
0.2	$\frac{1}{0.8} = 1.25$	\$1b × 1.25 = \$1.25b

Illustrating the impact of the spending multiplier

Using AD/AS analysis we can apply the spending multiplier to determine and illustrate the effect of an increase in aggregate expenditures in an economy experiencing a recessionary gap. Consider the model below.



The economy is producing \$1.5 trillion below its full employment level of output. Assume

the **MPC = 0.6**. With this information, we can determine the size of an increase in expenditures that would be needed to restore full employment.

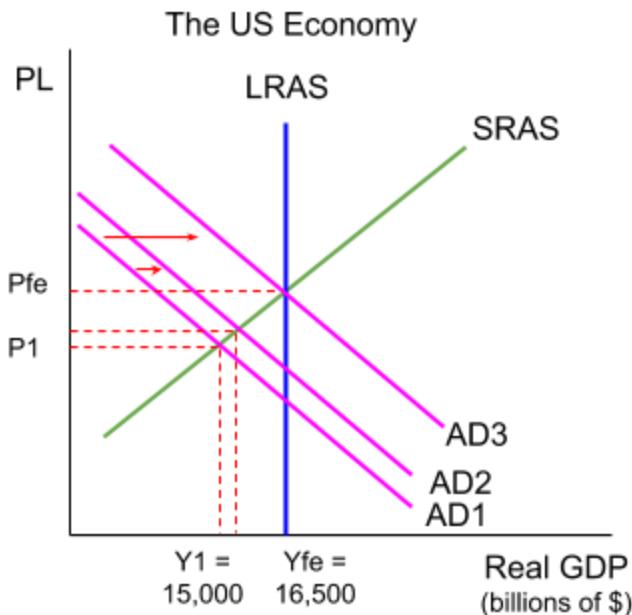
First let's calculate the size of the multiplier:

$$k = \frac{1}{1-MPC} = \frac{1}{0.4} = 2.5$$

To determine the needed increase in spending, we can divide the desired change in GDP by the spending multiplier.

$$\begin{aligned}\text{Needed change in spending} &= \frac{\text{desired } \Delta \text{ in GDP}}{\text{spending multiplier}} \\ &= \frac{\$1.5 \text{ trillion}}{2.5} \\ &= \$0.6 \text{ trillion } (\$600 \text{ billion})\end{aligned}$$

With a \$1.5 trillion recessionary gap and a spending multiplier of 2.5, government can expect a \$600 billion increase in spending to stimulate AD to a level that would result in the economy returning to full employment. The impact of the increase in aggregate spending and the multiplier are shown in the model below.



The \$600 billion increase in spending initially shifts AD from AD1 to AD2. As incomes increase because of the new spending, consumer spending increases and the initial change in spending is multiplied 2.5 times throughout the economy. Ultimately, AD shifts out to AD3, restoring full employment output at Yfe.

The multiplier effect works in both directions, of course. A fall in aggregate expenditures will also lead to a decrease in consumption, imports, taxation, and savings. Therefore, any

decrease in spending and incomes will be multiplied throughout the economy until the ultimate decrease in GDP will be greater than any initial decrease in aggregate expenditures.

2.3 The Macroeconomic Objectives

Introduction to the macroeconomic objectives

Macroeconomics studies aggregates of economic performance and how countries as a whole can achieve certain objectives. Policymakers, informed by economic data, models, and analysis, strive to implement policies aimed at achieving certain objectives, including:

- **Full employment:** When an economy is producing at its long-run potential level of output, it is achieving full employment. When producing at full employment, a country enjoys **low unemployment**, meaning that nearly everyone who is willing and able to work is contributing to the country's economic output.
- **Low inflation:** Economic stability means a country enjoys both low unemployment, but also low inflation. Price level instability, in the form of deflation or high inflation, creates an environment of economic uncertainty and can threaten a country's economy in many ways, leading to fluctuations in the business cycle and large swings in output and employment.
- **Economic growth:** Increasing both the country's actual output and its potential output over time assures that the average standard of living enjoyed by a country's residents will improve over time. Economic growth requires policies that promote improvements in human and physical capital, infrastructure, and other factors that contribute to rising living standards in a country.
- **Improved equity in the distribution of income:** In many countries, economic inequality poses a threat to social harmony and economic progress. The free market system, with all its efficiency and focus on productivity, tends to result in an inequitable distribution of opportunity, income, and wealth. Macroeconomic policy, therefore, must be designed to reduce the inequalities that arise in market economies.

A nation's policymakers, both its government and its central bank, have several policy tools at their disposal to promote the achievement of these objectives. The different policy options will be explored in detail in a later chapter. In this chapter, we will go through each of the four macroeconomic objectives in detail, and determine:

- The meaning of each objective,
- The measurement of each objective
- The graphical illustration of each objective

Objective #1 - Low unemployment

- Define the term unemployment.
- Explain how the unemployment rate is calculated.
- Explain the difficulties in measuring unemployment, including the existence of hidden unemployment, the existence of underemployment, and the fact that it is an average and therefore ignores regional, ethnic, age and gender disparities.

HL only objectives:

- Calculate the unemployment rate from a set of data.

Unemployment is defined simply as the state of being out of work, actively seeking work, but unable to get work. Note the following:

- Simply not having a job does not make an individual unemployed
- To be considered unemployed, an individual has to be of legal working age (over the age of 15 in most countries)
- Being underemployed is different than being unemployed. **Underemployment** refers to individuals who are working part-time but wish to work full time, or to people who are working in a job for which they are over-qualified.

Measuring unemployment

To determine the extent to which unemployment is a problem for an economy, economists calculate the **unemployment rate**, which is the percentage of the total labor force that is unemployed

$$\text{The Unemployment Rate (UR)} = \frac{\# \text{ of people unemployed}}{\# \text{ of people in the labor force}} \times 100$$

The **labor force** is the population of individuals in a nation who are of legal working age and are either employed or unemployed. Not included in the labor force are people who are:

- too young to work (0-14 years old)
- of working age but not working or looking for work
- retired
- of working age but in school full time
- institutionalized (in prison or jail)
- non-civilians (employees of the armed forces)

The labor force participation rate (LFPR)

The **labor force participation rate** is the percentage of the eligible population (15 or above, civilian, non-institutionalized) that is either employed or unemployed.

$$\text{Labor Force Participation Rate (LFPR)} = \frac{\# \text{ of employed and unemployed}}{\text{Total eligible population}} \times 100$$

Calculating the UR and the LFPR (HL only)

Consider the table below, which shows labor statistics for Brazil in 2010 and 2011.

Year	Eligible Population (millions of people 15 or above)	Number of people Employed (million)	Number of people Unemployed (millions)
2010	200	100	20
2011	220	105	22

From this data, we can calculate the following:

In 2010:

- The UR = $\frac{\# \text{ of people unemployed}}{\# \text{ of people in labor force}} \times 100 = \frac{20m}{120m} \times 100 = 16.67\%$
- The LFPR = $\frac{\# \text{ of people in labor force}}{\# \text{ of people in eligible population}} \times 100 = \frac{120m}{200m} \times 100 = 60\%$

In 2011

- The UR = $\frac{22m}{127m} \times 100 = 17.32\%$
- The LFPR = $\frac{127m}{220m} \times 100 = 57.72\%$

The LFPR fell over this period, indicating that of the eligible population, a smaller percentage was working or trying to find work in 2011.

The UR increased over this period, indicating that it was harder to find a job in 2011 than in 2010. Even though more people are employed, the percentage of the total labor force that is unemployed has increased.

Consequences of unemployment

- Discuss possible economic consequences of unemployment, including a loss of GDP, loss of tax revenue, increased cost of unemployment benefits, loss of income for individuals, and greater disparities in the distribution of income.
- Discuss possible personal and social consequences of unemployment, including increased crime rates, increased stress levels, increased indebtedness, homelessness and family breakdown.

Unemployment has several consequences for individuals, society, and the economy as a whole.

The Consequences of Unemployment	
For the Individual	<ul style="list-style-type: none"> • Decreased household income: reduces households' ability to buy the necessities and therefore reduces the standards of living of the unemployed • Increased levels of psychological and physical illness: Studies show that stress, depression, undernourishment, and other physical and mental effects arise from chronic unemployment
For Society	<ul style="list-style-type: none"> • Increased poverty and crime: There is a correlation between the level of unemployment in an economy and crime rates; the higher unemployment the more people will turn to crime to meet their basic needs • Transformation of traditional societies: Unemployment in rural areas of developing countries has risen as the global economy has changed the structures of these countries' economies, forcing traditional societies to adapt and in some cases dissolve as people seek work in modern industries.

For the Economy	<ul style="list-style-type: none"> ● Less total demand for goods and services: Households in which there are unemployed people earn less income and thus consume less, leading to less demand for goods and services in the economy. ● Under-utilization of resources: A nation with high levels of unemployment is not achieving its production possibilities, thus peoples' standards of living are less than what is possible. ● Downward pressure on wages for the employed: A large pool of unemployed workers increases the supply of available labor and thus reduces the wages offered to all workers.
------------------------	---

Types and causes of unemployment

- Describe, using examples, the meaning of frictional, structural, seasonal and cyclical (demand-deficient) unemployment.
- Distinguish between the causes of frictional, structural, seasonal and cyclical (demand-deficient) unemployment.
- Explain, using a diagram, that cyclical unemployment is caused by a fall in aggregate demand.
- Explain, using a diagram, that structural unemployment is caused by changes in the demand for particular labour skills, changes in the geographical location of industries, and labour market rigidities.
- Evaluate government policies to deal with the different types of unemployment.

Unemployment may take several forms, and arise from different macroeconomic conditions. The table below introduces the different types of unemployment and identifies their causes.

Type of Unemployment	Definition and Causes
Frictional Unemployment	<p>Frictional unemployment consists of those searching for jobs or waiting to take jobs soon; it is regarded as somewhat desirable, because it indicates that there is mobility in labor markets as people change or seek jobs.</p> <p>Seasonal unemployment is a type of frictional unemployment; it includes people who work seasonal jobs and are out of work for short periods of time between seasons.</p>
Structural Unemployment	<p>Structural unemployment arises due to changes in the structure the economy and a mismatch between the type of labor demanded and the type supplied, e.g. when certain skills become obsolete or geographic distribution of jobs changes. Examples: Glass blowers were replaced by bottle-making machines. Coal miners were displaced when natural gas began replacing coal for electricity generation. Airline mergers displaced many airline workers in 1980s. Foreign competition has led to downsizing in U.S. industry and loss of jobs.</p>

Cyclical Unemployment	Cyclical unemployment arises during the recession phase of the business cycle; sometimes called demand-deficient unemployment. Cyclical unemployment is caused by a fall in demand for the nation's output, which causes a loss of jobs in the economy.
------------------------------	---

The natural rate of unemployment (NRU)

When an economy is producing at its full employment level it experiences only frictional and structural unemployment. The unemployment that exists when an economy is producing at full employment is called the **natural rate of unemployment**.

$$\text{Natural rate of unemployment (NRU)} = \text{frictional unemployment} + \text{structural unemployment}$$

The deviation of the actual unemployment rate from the natural rate of unemployment is known as **cyclical unemployment**.

- If a country is in a recession, there will be positive cyclical unemployment; people have lost their jobs due to weak demand for the country's output.
- If a country is overheating, the actual unemployment rate will be lower than the natural rate; demand for the country's output is high and workers who would normally be structurally or frictionally unemployed are instead employed.

Causes of structural unemployment

Structural unemployment arises due to changing technology or other factors that result in a mismatch between the skills of a nation's workforce and the needs of employers.

- If the technology used in production changes and becomes more capital intensive, the demand for workers who were previously needed to produce goods will decline
- If foreign countries can produce goods more cheaply, then domestic demand for certain types of labor will fall.
- If a nation's education and jobs training system does not prepare workers with the skills demanded in the labor market, structural unemployment will rise over time.

Causes of frictional unemployment

If unemployed workers cannot quickly and easily be matched up with firms that demand labor, then **frictional unemployment** will be higher and last longer than if it is easy for employers and potential employees to find one another

Causes of cyclical unemployment

Cyclical unemployment is caused by a reduction in output to a level below the full employment level, also known as a recession. Recessions could be caused by either a decrease in aggregate demand or a decrease in short-run aggregate supply.

- If the total demand for a nation's output falls, firms will, in the short-run, reduce the number of workers they employ and reduce their output.
- If the costs of production faced by a nation's producers suddenly rise, firms will employ fewer workers to try and remain profitable. National output falls and

unemployment rises.

Changes in the natural rate of unemployment

A country's **natural rate of unemployment (NRU)** can change over time if there is a change in the level of frictional or structural unemployment. For example:

- An improvement in the skills and knowledge of the country's labor force (also known as "human capital") can reduce the level of structural unemployment and lower a country's NRU
- New technologies (such as job-search websites) that match job-seekers with potential employers will reduce the level of frictional unemployment and reduce the NRU

Graphical analysis of structural unemployment

To illustrate structural unemployment, a microeconomic diagram known as a labor market graph is needed. Structural unemployment arises when demand for particular types of labor (workers with certain skill sets) falls. For example, assume new robotic manufacturing technologies are making factory workers increasingly obsolete. Demand for factory workers will fall, as in the graph below.



Observe from the graph above:

- Due to improved manufacturing technology, demand for factory workers has fallen from D1 to D2.
- The wage rate (WR1) does not decrease due to "sticky wages" - labor unions, contracts, unemployment benefits, minimum wage laws, and other labor market rigidities prevent the equilibrium wage rate from falling to the new intersection of supply and demand.

- At WR₁ more workers are seeking jobs in factories (Q_s) than there are factory owners willing to hire (Q_d)
- The “excess supply of labor” from $Q_d - Q_s$ represents the structural unemployment resulting from improvements in robotic manufacturing technology.

Policies to reduce structural unemployment

Notice that one solution to reducing structural unemployment is increasing labor market flexibility. If the wage rate were able to adjust to the fall in demand for labor more easily, factory owners would be willing to hire more factory workers, choosing not to replace them with robots.

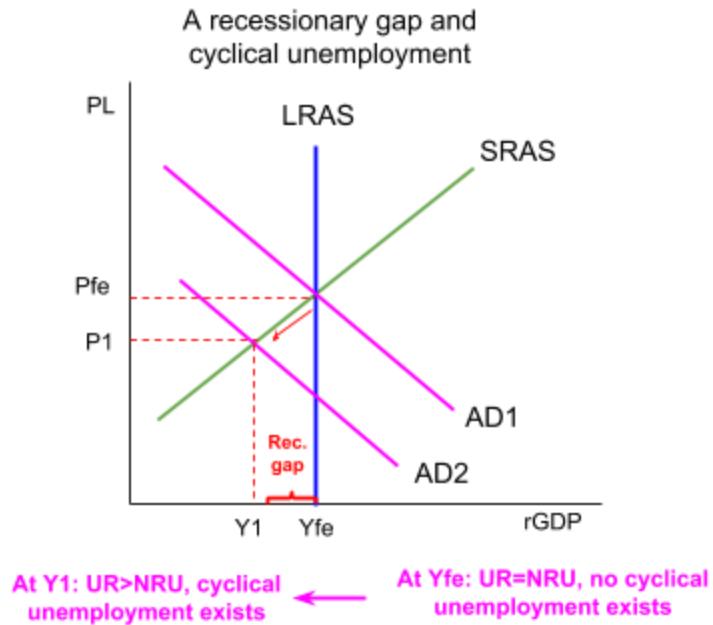
On the other hand, the lower wage rate would lead fewer people to seek jobs as factory workers, reducing the quantity supplied in the labor market. A greater quantity demanded and smaller quantity supplied would restore equilibrium and eliminate structural unemployment.

In a later chapter we'll explore “supply-side policies” that would increase labor market flexibility and help reduce structural unemployment. Some other government policies to deal with structural unemployment include:

- Investing in job and skills training programs to match the skills possessed by workers with those demanded by employers.
- Investing in education so that people entering the labor force are equipped with the skills demanded by the country's employers.

Graphical analysis of cyclical unemployment

Cyclical unemployment arises due to a fall in aggregate demand and the accompanying recession phase of the business cycle. Therefore, graphing cyclical unemployment requires us to use an AD/AS diagram and show a fall in AD and a recessionary gap, as seen below.



Observe from the graph:

- As AD falls from AD₁ to AD₂ equilibrium national output falls from Y_{fe} to Y₁.
- At a lower level of output, fewer workers are needed.
- Firms will reduce employment in order to cut costs and reduce output in the face of lower demand for their products.
- Cyclical unemployment results from a fall in aggregate demand.

One major and important difference between cyclical and structural unemployment is that cyclically unemployed workers DO possess skills that would be in demand if only the economy were producing at its full employment level. In contrast, structurally unemployed workers do not possess relevant skills for the country's economy.

In other words, when cyclical unemployment exists, it is evidence that the economy as a whole is “sick”, or in recession. When structural unemployment exists, it is evidence that certain workers in the country are in need of retraining to equip them with skills more relevant to the country’s economy.

Policies to reduce cyclical unemployment

From the graph we can observe a couple of options policymakers have for reducing cyclical unemployment:

- **Demand-side policies:** Policies aimed at increasing AD will reduce the level of cyclical unemployment in a country. These include **expansionary fiscal policies** and **expansionary monetary policies**; both will be explored in more detail in a later chapter.
- **Supply-side policies:** Policies aimed at increasing SRAS will reduce the level of unemployment. These include **interventionist supply-side policies** and **market-based supply-side policies**; both will be explored in more detail in a later

chapter.

Policies to reduce frictional unemployment

Frictional unemployment is a completely natural phenomenon and will never be eliminated, nor should it be. People must always enter the labor force for the first time and seek jobs; such entrance into the labor force, either through graduation from university or high school, immigration, or people voluntarily out of the labor force reentering it, is essential for an economy.

However, means of reducing the period of time people are frictionally unemployed can improve the economy's efficiency by helping employers find the employees they need more quickly. Measures for reducing frictional unemployment include:

- Job placement agencies
- Hiring/recruiting services
- Websites to help workers seeking jobs get their resumes out to potential employers

Mobile and web technologies have vastly improved the ability of potential workers to find potential employers, reducing the amount of frictional unemployment, and it could therefore be argued, reducing the natural rate of unemployment.

Objective #2 - Low inflation: The meaning of inflation, disinflation and deflation

- Distinguish between inflation, disinflation and deflation.
- Explain that inflation and deflation are typically measured by calculating a consumer price index (CPI), which measures the change in prices of a basket of goods and services consumed by the average household.

HL only objectives:

- Construct a weighted price index, using a set of data provided.
- Calculate the inflation rate from a set of data.

The meaning of inflation

Inflation is defined as an increase in the average price level of goods and services in a nation over time.

The percentage change in the price level between two periods of time is known as the **inflation rate**. The inflation rate can be positive or negative. Some key terms relating to the measurement of the price level are:

- **Inflation:** When the price level increases between two periods of time (a positive inflation rate).
- **Disinflation:** When the rate of inflation decreases between two periods of time (a decrease in the inflation rate).
- **Deflation:** When the price level decreases between two periods of time (a negative inflation rate).

Measuring inflation

To determine whether a nation's price level is increasing or decreasing over a particular time period, economists use what is known as a price index.

- **The consumer price index (CPI)** measures the price of a set basket of consumer goods (usually includes hundreds or even thousands of goods that the typical household in a nation consume) between one time period and another.
- **The inflation rate** is the percentage change in the CPI between two years:

$$\text{The inflation rate} = \frac{\text{CPI in year 2} - \text{CPI in year 1}}{\text{CPI in year 1}} \times 100$$

Calculating inflation using a CPI (HL only)

The CPI for a particular year is the price of a basket of goods in that year divided by the price of the same basket in a base year.

To calculate inflation between two years, we first must determine the CPIs for the two years in question. Assume the CPI is made up of just three goods, whose prices during two years are indicated in the table below.

Good or service	Price in 2018	Price in 2019
Pizza	10€	10.50€
Haircuts	20€	19€
Wine	8€	10€
Total basket price	38€	39.50€

Assume 2018 is the base year, and we want to calculate inflation between 2018 and 2019.

First we must calculate the price indices for the two years:

$$\text{CPI for 2018} = \frac{\text{Price of the basket of goods in 2018}}{\text{Price of the basket in base year}} \times 100 = \frac{38}{38} \times 100 = \mathbf{100}$$

$$\text{CPI for 2019} = \frac{\text{Price of the basket of goods in 2019}}{\text{Price of the basket in base year}} = \frac{39.5}{38} \times 100 = \mathbf{103.9}$$

The CPI for 2018 is 100. Since 2018 is our base year, we are comparing the price of the goods in that year to itself, so of course the index is equal to 100.

The price of the same basket in 2019 is €39.50, which when divided by the base year price and multiplied by 100 results in a CPI of 103.9

With the CPIs known, we can calculate the rate of inflation:

$$\text{The inflation rate} = \frac{\text{CPI in year 2} - \text{CPI in year 1}}{\text{CPI in year 1}} \times 100 = \frac{103.9 - 100}{100} \times 100 = \mathbf{3.9\%}$$

The CPI increased by 3.9% between 2018 and 2019, indicating that inflation equaled 3.9%.

Using a weighted price index to calculate inflation (HL only)

Because not all the goods measured in a nation's CPI are equally important to the typical household, governments assign weights to particular categories of goods.

- For example, food and beverages make up approximately 15% of the typical household budget in a given year. But housing (either rental payments or mortgage payments) make up 40%.
- In this example, housing prices should be weighted more heavily than food and beverages

Consider the table showing the prices of the three goods measured in a CPI in two years, including the weight given to each good based on the percentage of the typical consumer's income spent on it.

Good	Price in 2018	Price in 2019	Weight
Banana	\$2	\$1.50	25%
Haircut	\$11	\$10	30%
Taxi ride	\$8	\$10	45%
		Total	100%

To establish a price index with 2018 as the base year, we must calculate the weighted price of the basket of goods for 2018. To do this, we multiply the average price of each good by its weight, expressed in hundredths.

2018:

- Banana = $2 \times 0.25 = 0.5$
- Haircut = $11 \times 0.3 = 3.3$
- Taxi ride = $8 \times 0.45 = 3.6$
- $0.5+3.3+3.6 = 7.7$

Weighted price index for 2018 = **7.7**

2019:

- Banana = $1.5 \times 0.25 = 0.375$
- Haircut = $10 \times 0.3 = 3$
- Taxi ride = $12 \times 0.45 = 5.4$
- $0.375+3+5.4 = 8.775$

Weighted price index for 2019 = **8.775**

The prices of the goods in the CPI have now been adjusted for their relative importance to the consumer. A change in the price taxi rides (weighted at 45%) will now have a larger impact on the overall inflation rate than a change in the price of bananas (which are only weighted at 25%).

With the weighted price indices determined, we can calculate inflation between 2018 and 2019:

$$\text{Inflation rate} = \frac{8.775 - 7.7}{7.7} \times 100 = \mathbf{14\%}$$

Degrees of inflation

A country's CPI does not always increase at a steady rate. The CPI could fall over time (as goods get cheaper), it could increase very slowly (low inflation), or very rapidly (high inflation). The table below distinguishes between different degrees of inflation.

Degrees of Inflation, from low to high	
Deflation:	Deflation refers to a decrease in the average price level of goods/services over time. <ul style="list-style-type: none"> If the CPI for one year is smaller than the CPI from a previous year, then the inflation rate will be negative. Deflation is considered highly undesirable because it discourages investment and consumption (households and firms prefer to postpone spending until prices are lower in the future) and therefore can lead to recession and rising unemployment.
Low inflation:	Inflation rates of between 0-5% are considered low and stable <ul style="list-style-type: none"> This is the desired range for most countries, over which consumers' confidence over the stability of future prices is sound; businesses and households can invest, spend and save without fear of future erosions in the values of their savings and investments.
High inflation:	Inflation rates of greater than 5% are considered high in most countries <ul style="list-style-type: none"> At high inflation rates, firms and households will rush to spend their money now before its value is eroded by higher prices. The race to spend while money is dear causes AD to grow rapidly, causing demand-pull inflation, reducing real incomes and contributing to instability across the economy

Shortcomings of the inflation rate as a macroeconomic measure

- Explain that different income earners may experience a different rate of inflation when their pattern of consumption is not accurately reflected by the CPI.
- Explain that inflation figures may not accurately reflect changes in consumption patterns and the quality of the products purchased.
- Explain that economists measure a core/underlying rate of inflation to eliminate the effect of sudden swings in the prices of food and oil, for example.
- Explain that a producer price index measuring changes in the prices of factors of

production may be useful in predicting future inflation.

Effect on different income earners

The inflation rate, like other macroeconomic measurements, is an **aggregate measure**, in this case of what the typical household consumes. However, not all households' consumption patterns will be accurately reflected by the CPI.

For example, a lower income household may be more affected by inflation in their everyday consumption decisions than a wealthier households. Poorer consumers tend to spend a greater percentage of their income on goods and services, whereas richer consumers tend to save more. Therefore, an increase in the CPI could have a disproportionate effect on lower income consumers than on the rich, who are saving more of their income.

Additionally, the composition of the basket of goods consumed by different households will be very different. Higher income households may be able to substitute imported goods for domestically produced goods when domestic inflation accelerates, sheltering them from rising prices at home.

The inability of the CPI to reflect the consumption patterns of ALL households is one obvious shortcoming of its usefulness as a measurement of economic well-being.

Changing consumption patterns and product quality

The goods included in CPI may remain fixed over several years, even as the goods purchased by actual consumers change over time. If the goods in a CPI are not updated frequently, the usefulness of the inflation rate will diminish as consumers switch to newer, different products over time.

Additionally, the CPI might overstate inflation as prices rise even as the quality of the goods consumers are buying increases exponentially. For example, a \$25,000 car (say, the Toyota Camry) purchased in 2008 might have increased to \$30,000 by 2018, a 20% increase in price (an average of 2% per year over 10 years).

However, the quality of the car purchased in 2018 may have increased exponentially. Improvements in safety, comfort, entertainment and navigation systems, self-driving or "autopilot" capabilities, and other factors affecting the consumer's experience of the 2018 Camry are not reflected in that average 2% increase in its price over 10 years.

When technology and product quality increase disproportionately compared to goods' prices, the inflation rate will overstate the burden of higher prices on consumers.

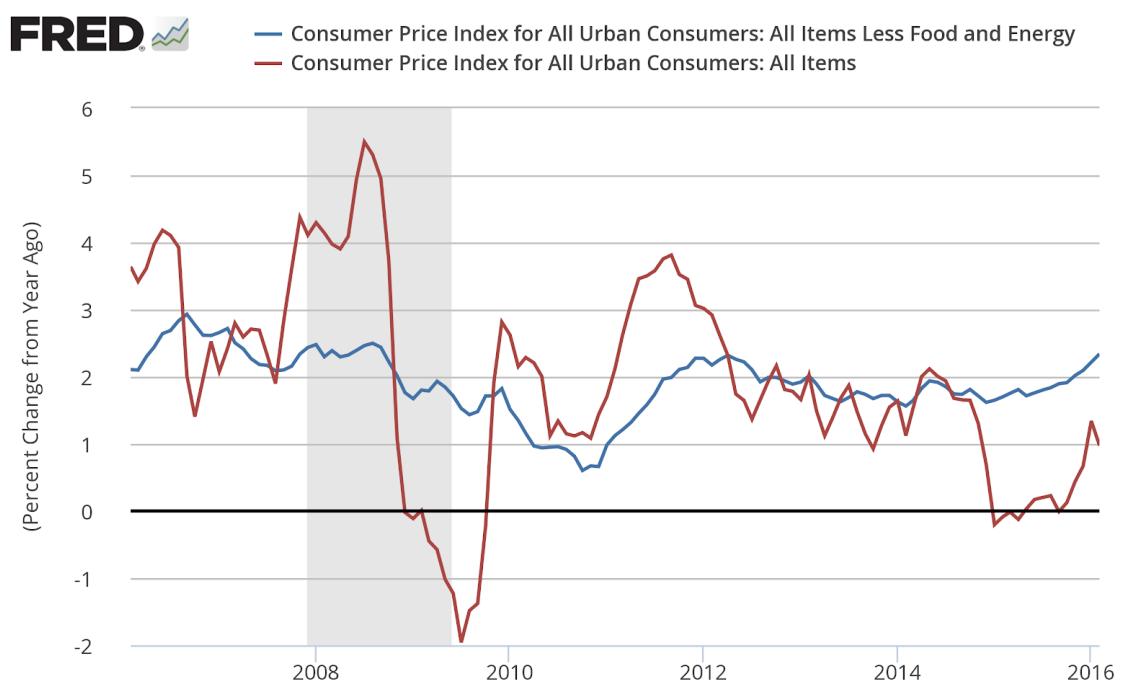
Core inflation

A country's inflation rate could appear rather volatile when the prices of certain products within the CPI fluctuate wildly in a certain year. Energy and food prices tend to be rather volatile, due to the highly inelastic supply of such commodities in the short-run.

For example, during the year when oil prices fell from their 2008 high of nearly \$150 per barrel to \$40 per barrel, the CPI increased less than it would have had oil prices remained high.

In order to provide a picture for how MOST goods and services are changing in price, not including those commodities whose prices regularly fluctuate, affecting the CPI, economists measure what is known as the **core CPI**, to determine **core inflation**. Core inflation measures the price of a basket of goods consumed by the typical household, excluding energy and food. It gives a clearer picture of the long-run trends in living costs, ignoring short-run fluctuations in certain commodity prices.

The chart below compares the United States' CPI for all items (in red) to the core CPI (in blue) between 2007 and 2017. The area shaded in gray indicates the Great REcession of 2008-2009.



Observe from the chart:

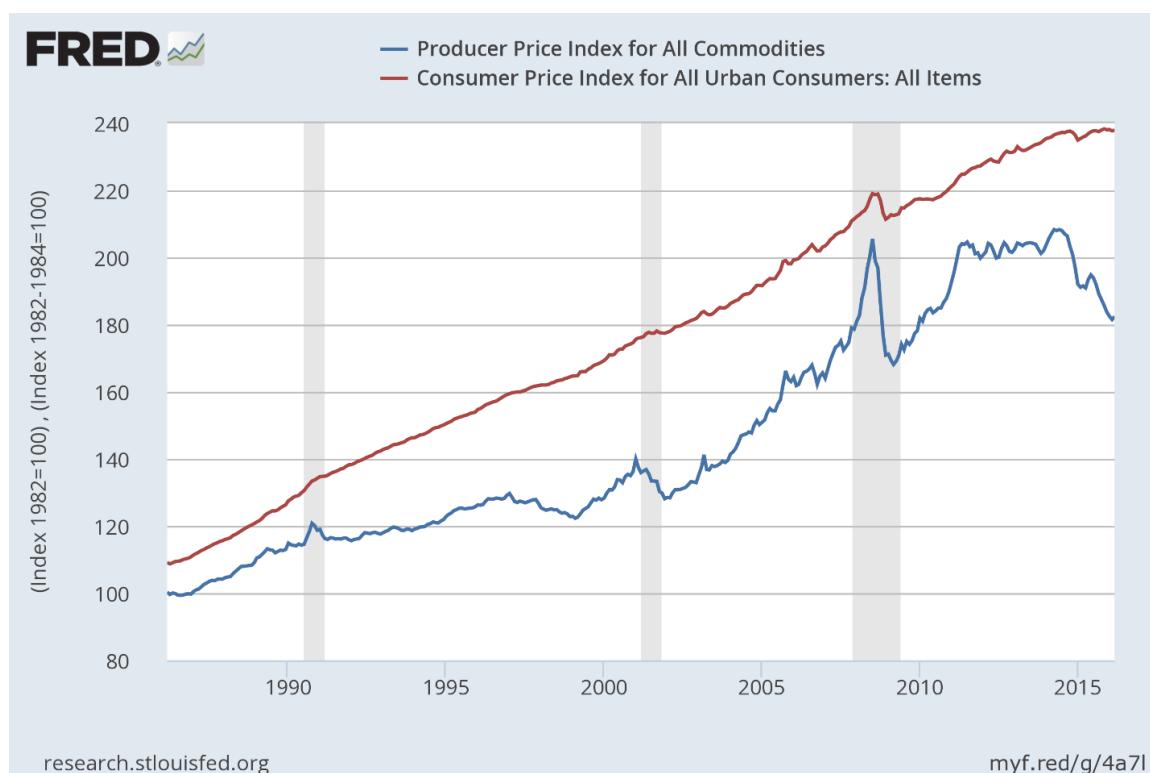
- The “headline” inflation rate (in red) fluctuates more than the “core” inflation rate (in blue)
- Both headline and core inflation rates fell during the Great Recessions, due to lower aggregate demand in the economy as a whole.
- The headline inflation rate became negative (deflation) in 2009, due to the fact that commodity prices (such as food and energy) fell more than most consumer prices during the recession.

Alternative measures of inflation - the Producer Price Index

The CPI measures price changes in goods consumed by the typical household; therefore, it does not represent how prices that affect firms are changing over time.

The **Producer Price Index (PPI)** measures the price of a basket of goods purchased by the typical producer in a country over time. The PPI includes more primary commodities, including mining, manufacturing, agriculture, fishing, and forestry - as well as natural gas, electricity, construction, and other goods purchased by a country's business firms.

The chart below compares the United States' CPI and PPI from 1985 through 2015.



Observe from the chart:

- The CPI and PPI generally rise together over time.
- During recessions (gray periods in the chart), the PPI tends to experience steep declines, while the CPI typically experiences slower growth (disinflation) or a relatively small drop (mild deflation).
- The PPI (in blue) fluctuates more than the CPI, since commodity prices are more volatile than prices for finished goods (due to their relatively inelastic supply and demand).

Consequences of inflation

- Discuss the possible consequences of a high inflation rate, including greater uncertainty, redistributive effects, less saving, and the damage to export

competitiveness.

High inflation, like high unemployment, has several negative effects on households, firms and the overall economy.

The Consequences of High Inflation	
Lower Real Incomes	A households' real income is its nominal income adjusted for any inflation in the economy. The more prices rise, the less a certain amount of income can buy for households. Higher inflation makes consumers feel poorer, since the real value of their incomes falls when inflation rises.
Lower Real Interest Rates for Savers	The real interest rate is the nominal interest rate minus the inflation rate. For example, if you have a savings account offering a 5% interest rate, and inflation is 2%, the real return on your savings is only 3%. But if inflation increases to 4%, your real return is just 1%. Inflation reduces the incentive for households to save, driving up current consumption, causing even more inflation.
Higher nominal interest rates for borrowers	When banks anticipate high inflation in the future, they will raise the interest rates they charge borrowers today. This increases the cost of borrowing money to invest in new capital or to buy homes or expensive durable goods.
Lower real interest rates for borrowers	Inflation reduces the real interest rate for borrowers. The money paid back by borrowers is worth less than the money borrowed when there is inflation, thus the real interest paid is lower. Example: If a borrower faces a 5% interest rate on a loan, and the inflation rate increases from 2% to 3%, the real interest owed decreases from 3% to 2%.
Reduced debt burden for debtors	Similar to above, inflation erodes the real value of an individual's or a nation's debt. The value of the money owed by a debtor decreases as inflation increases.
Reduced international competitiveness	A country experiencing high inflation will find demand for its goods fall among international consumers, as they become more expensive compared to other countries' goods. Also, higher prices and wages will reduce foreign investment in the country as firms do not wish to produce where costs are rising, rather where costs will be low in the future.

Consequences of deflation

- Discuss the possible consequences of deflation, including high levels of cyclical unemployment and bankruptcies.

Deflation, a decrease in the average price level, sounds like a good thing. But it is not, and in some circumstances can be worse for an economy than mild inflation.

The Consequences of Deflation	
Rising Unemployment:	With the expectation of lower future prices for their output, and with low demand for goods and services, firms are likely to lay off workers, leading to higher unemployment and downward pressure on workers' wages across the economy
Delayed consumption:	With the expectation of future price decreases, households will increase savings and decrease spending. The decrease in current consumption can lead to further deflation and contribute to a deflationary spiral, in which lower prices lead to lower AD which leads to even lower prices
Declining investment:	If firms expect less demand for their output in the future, they'll invest less now, which could result in slower economic growth, as the nation's capital stock depreciates over time and is not being replenished at a rate that will promise sustained growth
Cost to borrowers:	Deflation causes the value of money to increase over time. Therefore, the real debt burden of borrowers increases as the price level falls. Bankruptcies result as borrower's incomes fall while the value of the money they must pay back increases.

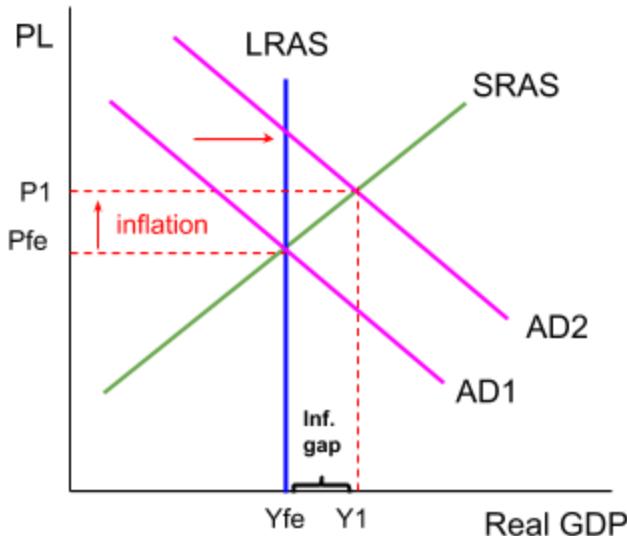
Types and causes of inflation

- Explain, using a diagram, that demand-pull inflation is caused by changes in the determinants of AD, resulting in an increase in AD.
- Explain, using a diagram, that cost-push inflation is caused by an increase in the costs of factors of production, resulting in a decrease in SRAS.
- Evaluate government policies to deal with the different types of inflation.

Inflation can be caused by one of two ways, either as a result of an increase in aggregate demand or as a result of a decrease in aggregate supply.

Demand-pull inflation occurs when there is an increase in total demand for a nation's output, either from domestic households, foreign consumers, the government or firms (C, Xn, G or I). When demand increases without a corresponding increase in aggregate supply, the nation's output cannot keep up with the demand, and prices are driven up as goods become scarcer.

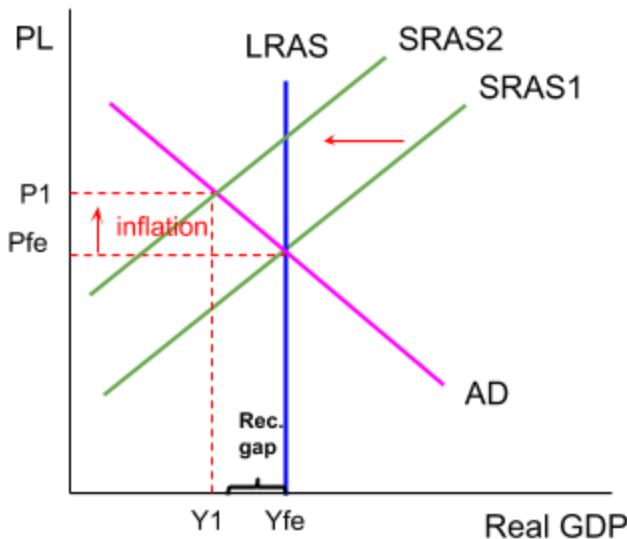
Demand-pull inflation can be illustrated in the AD/AS model:



Cost-push inflation occurs as the result of a negative supply shock, arising from a sudden, often unanticipated, increase in the costs of production for the nation's producers. Cost-push inflation could result from any of the following:

- Increase in the wage rate
- Increase in resource costs
- Increased energy or transportation costs
- Increased regulation by the government
- Increased business taxes
- Reduction in the exchange rate

Cost-push inflation can be illustrated in the AD/AS model:



Government policies to deal with inflation

Depending on its cause, policymakers have different options for dealing with inflation.

Demand-pull inflation requires the use of **demand-side policies**, such as **contractionary fiscal policy** or **contractionary monetary policy**.

- Reducing government spending or increasing taxes will reduce aggregate demand and reduce the rate of inflation
- Increasing interest rates (an action taken by the central bank) reduces the level of investment and consumption and reduces the inflation rate.

Demand-side policies will be explored further in a later chapter.

Cost-push inflation requires the use of **supply-side policies**, including **interventionist** and **market-based supply-side policies**, including.

- Labor market reforms
- Trade liberalization
- Investment in human capital, infrastructure, and technology,
- Anti-monopoly regulations
- Reductions in minimum wage and unemployment compensation.

Supply-side policies will be explored further in a later chapter.

Possible relationships between unemployment and inflation (HL only)

- Discuss, using a short-run Phillips curve diagram, the view that there is a possible trade-off between the unemployment rate and the inflation rate in the short run.
- Explain, using a diagram, that the short-run Phillips curve may shift outwards, resulting in stagflation (caused by a decrease in SRAS due to factors including supply shocks).
- Discuss, using a diagram, the view that there is a long-run Phillips curve that is vertical at the natural rate of unemployment and therefore there is no trade-off between the unemployment rate and the inflation rate in the long run.
- Explain that the natural rate of unemployment is the rate of unemployment that exists when the economy is producing at the full employment level of output.

The short-run Phillips curve (SRPC)

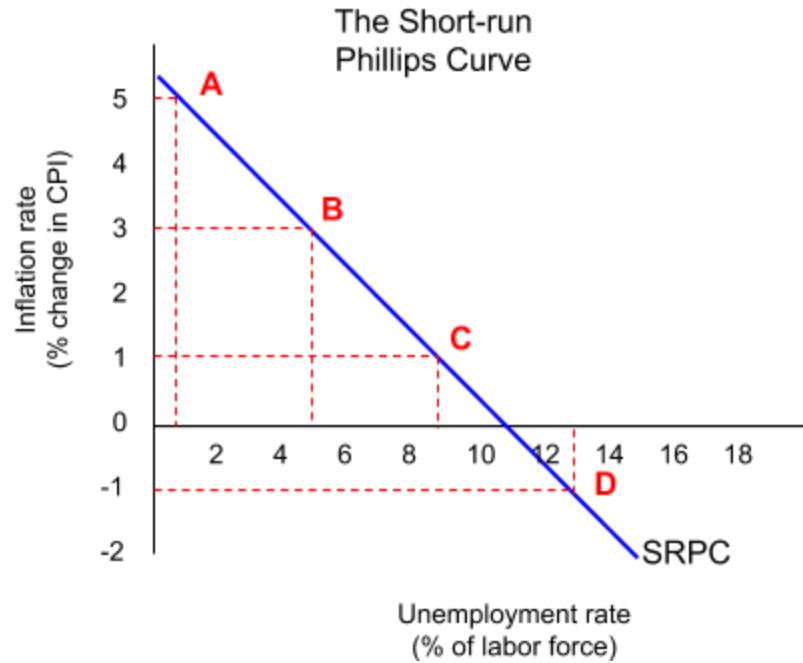
As we have seen in our AD-AS analysis, whenever a country experiences a shift in aggregate demand, in the short run there is a change in both inflation and unemployment.

- When AD increases, unemployment falls and inflation increases.
- When AD decreases, unemployment increases and inflation falls.

The short-run tradeoff between inflation and unemployment is so fundamental to macroeconomics that economists have developed a model just to illustrate this relationship.

The Phillips curve model shows the rate of inflation on the vertical axis and the unemployment rate on the horizontal axes. The curve itself is downward sloping, showing the inverse relationship between the two in the short run.

Assume that in the country represented by the Phillips Curve below the **natural rate of unemployment** is 5% and the **target inflation rate** is between 2% and 3%.



Points along a country's SRPC correspond with different levels of aggregate demand the country could experience at any given period of time:

- At **point A** aggregate demand is relatively high, resulting in a high inflation rate (5%) and a very low unemployment rate (<1%). The economy is producing beyond full employment with a positive output gap.
- At **point B** AD has fallen to a level around full employment. Inflation is within the target range of 2%-3% and unemployment is 5%, close to the natural rate of unemployment.
- At **point C** AD has fallen further and the economy has lower than desired inflation of 1% and unemployment that is greater than the natural rate at 9%. The economy has a recessionary gap.
- At **point D** AD has fallen far below full employment and the economy has inflation of -1% (deflation) and unemployment of 13%. The economy is experiencing a deep recession.

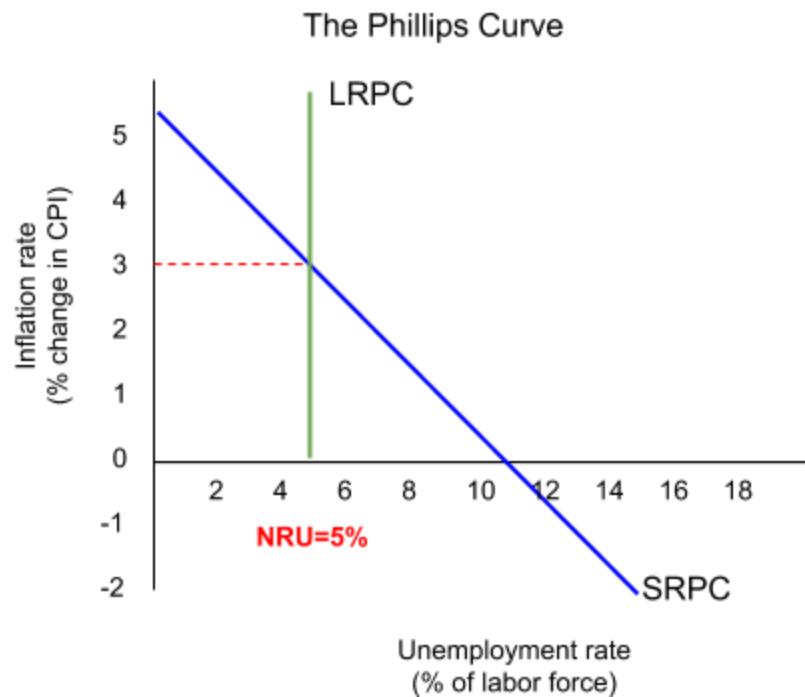
The long-run Phillips curve (LRPC)

The long-run relationship between inflation and unemployment can be illustrated by the long-run Phillips curve (LRPC), which is vertical at the natural rate of unemployment.

Recall from our study of the AD-AS model that the long-run aggregate supply (LRAS) curve is vertical at the full employment level of output, explained by the fact that wages and other input costs are fully flexible in the long run. Since output always returns to the full employment level in the long run, the unemployment rate will always return to the natural

rate of unemployment (NRU).

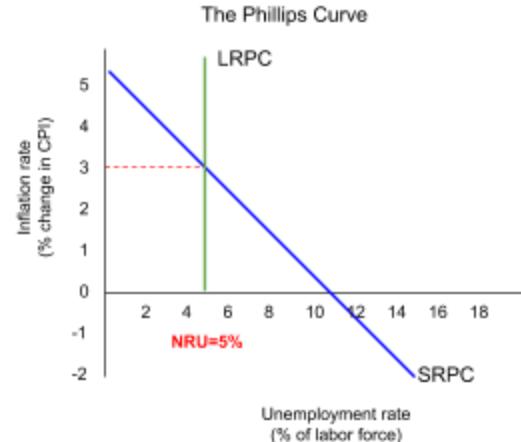
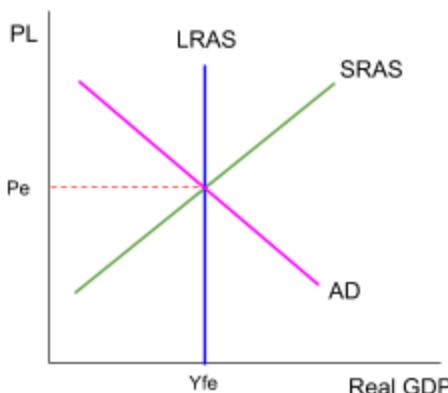
The graph below shows both the SRPC and the LRPC, which is vertical at the country's NRU of 5%.



Long-run equilibrium corresponds to the intersection of the SRPC and the LRPC. In the short run, when wages and other input prices are fixed, an economy can be producing anywhere along its SRPC. However, in the long run, when all prices are flexible, a country will return to its NRU.

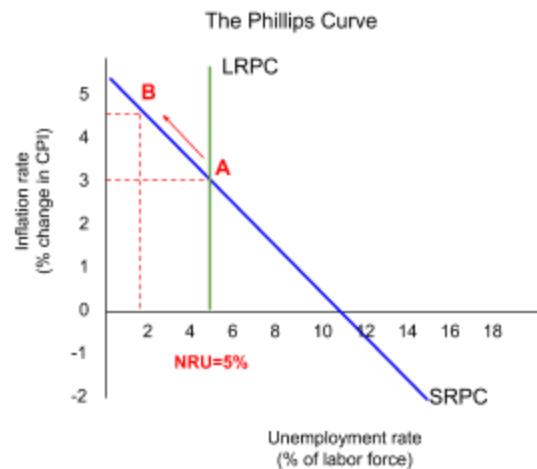
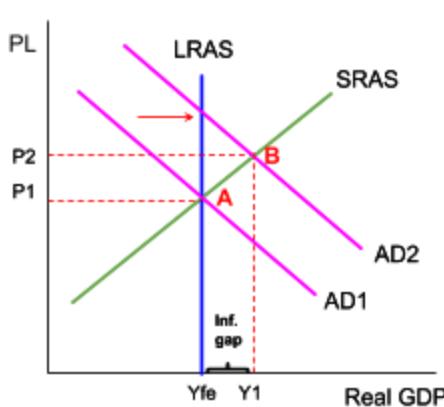
Demand shocks in the Phillips curve model

Changes to aggregate demand cause a movement along a country's SRPC. For example, assume a country is currently producing at its full employment level and is in its long-run equilibrium in both the AD-AS model and the Phillips Curve model.



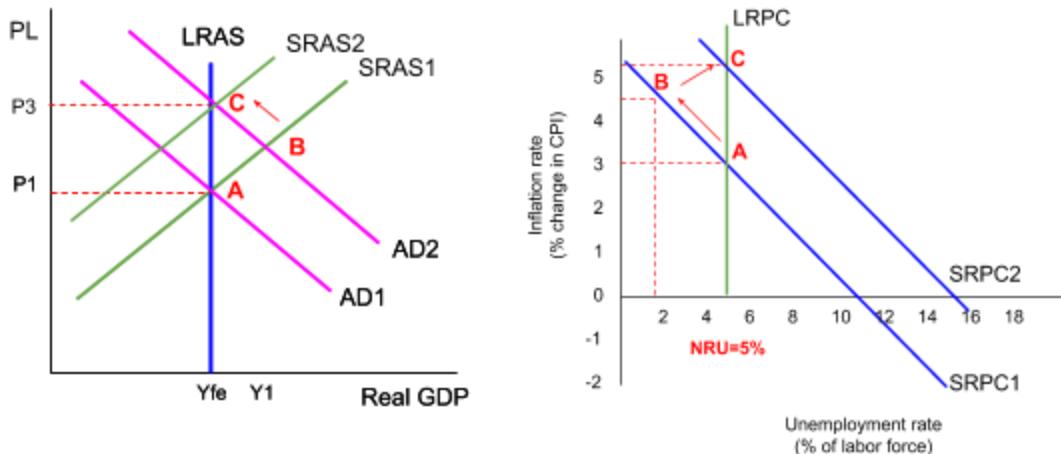
Positive demand shocks

A positive demand shock will cause an increase in output, employment, and the price level in the AD-AS model, and a movement up and to the left along the SRPC in the Phillips Curve model:



A move from point A to B in the AD-AS model causes a move from point A to point B in the Phillips Curve model.

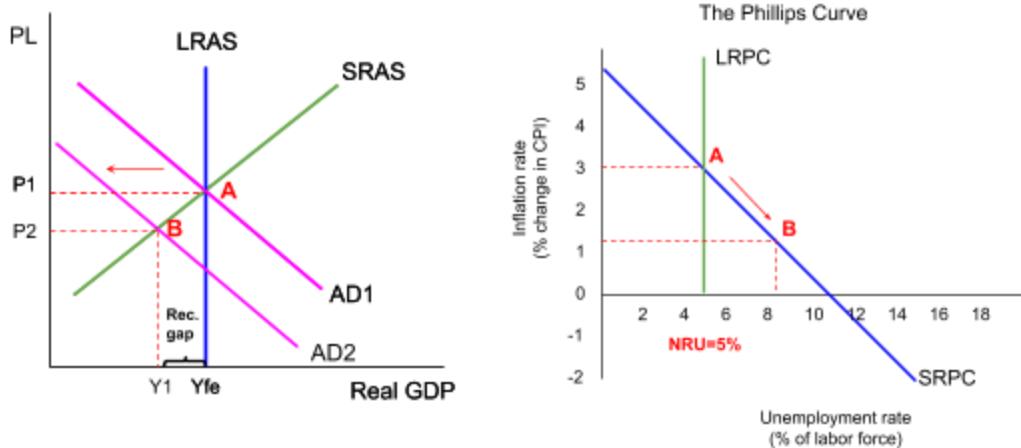
In the long run, an economy producing at point B (beyond full employment) will experience rising wages and input costs, causing the SRAS to decrease and output to return to the full employment level. As this happens, inflation will increase and unemployment will return to the NRU in the Phillips Curve model.



Rising wages and other input prices causes the SRAS to shift in, restoring full employment in the AD-AS model at a higher price level. The inward shift of the SRAS causes an outward shift of the SRPC, restoring the NRU in the Phillips curve model at a higher inflation rate.

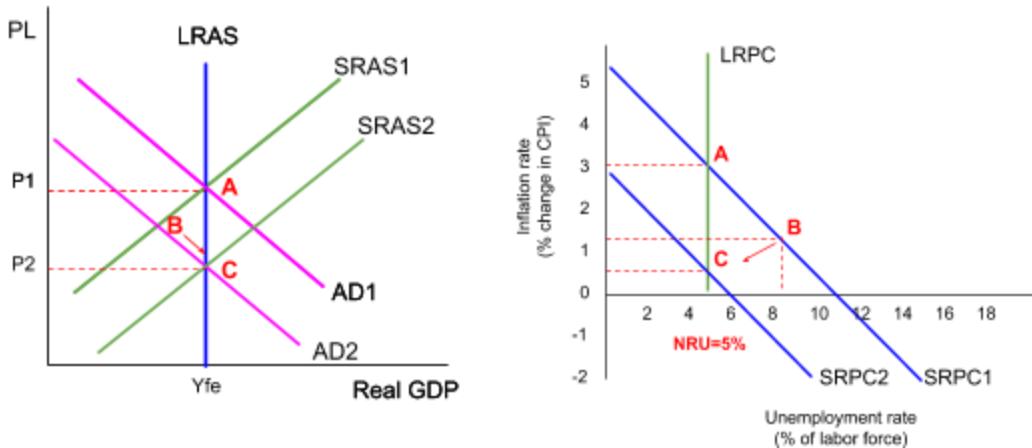
Negative demand shocks

A negative demand shock will cause a decrease in output, employment, and the price level in the AD-AS model, and a movement down and to the right along the SRPC in the Phillips Curve model:



A move from point A to B in the AD-AS model causes a move from point A to point B in the Phillips Curve model.

In the long run, an economy producing at point B (below full employment) will experience falling wages and input costs, causing the SRAS to increase and output to return to the full employment level. As this happens, inflation will decrease and unemployment will return to the NRU in the Phillips Curve model.

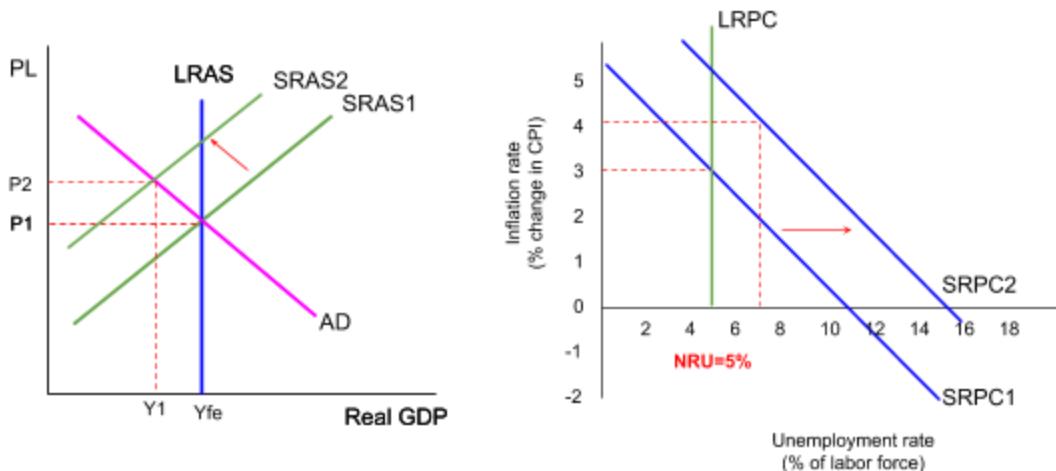


Falling wages and other input prices causes the SRAS to shift out, restoring full employment in the AD-AS model at a lower price level. The outward shift of the SRAS causes an inward shift of the SRPC, restoring the NRU in the Phillips curve model at a lower inflation rate.

Supply shocks in the Phillips curve model

Whenever a factor leads to a shifts in the short-run aggregate supply (SRAS) curve, the short-run Phillips curve (SRPC) shifts in the opposite direction.

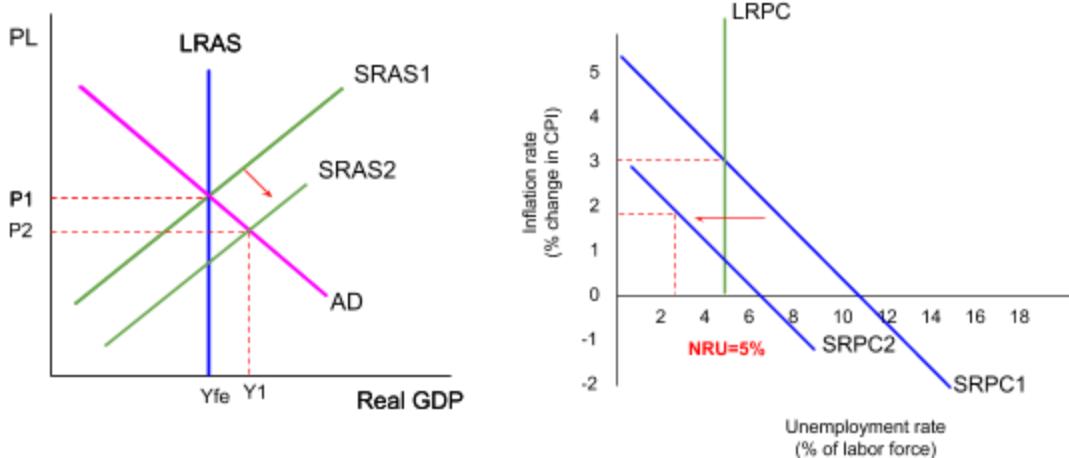
A negative supply shock causes both higher inflation and higher unemployment. For example, assume there is an unexpected increase in energy prices. SRAS will shift in as the cost of producing output increases. Firms reduce both employment and output, while raising prices to consumers.



A negative supply shock causes higher inflation and higher unemployment. A country experiences **stagflation** when SRAS shifts in and the SRPC shifts out. Stagflation is a mash-up of the words “stagnant” and “inflation”. In other words, the country’s economy stagnates while inflation increases.

A positive supply shock leads to more output and employment and lower prices. The SRPC

shifts inward, allowing a country to enjoy both lower inflation and lower unemployment.



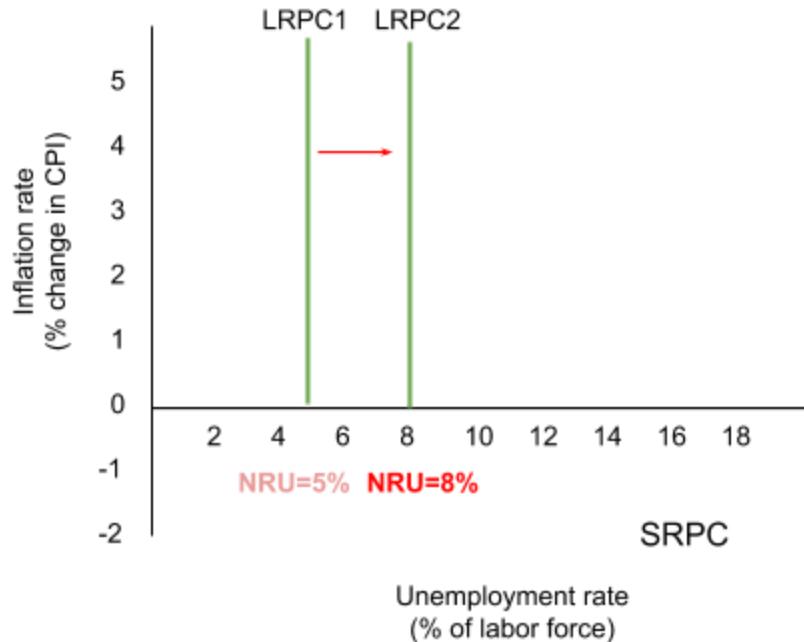
A positive supply shock causes output to increase and the price level to fall. More output means lower unemployment and a lower price level means lower inflation.

Shifts of the long-run Phillips curve

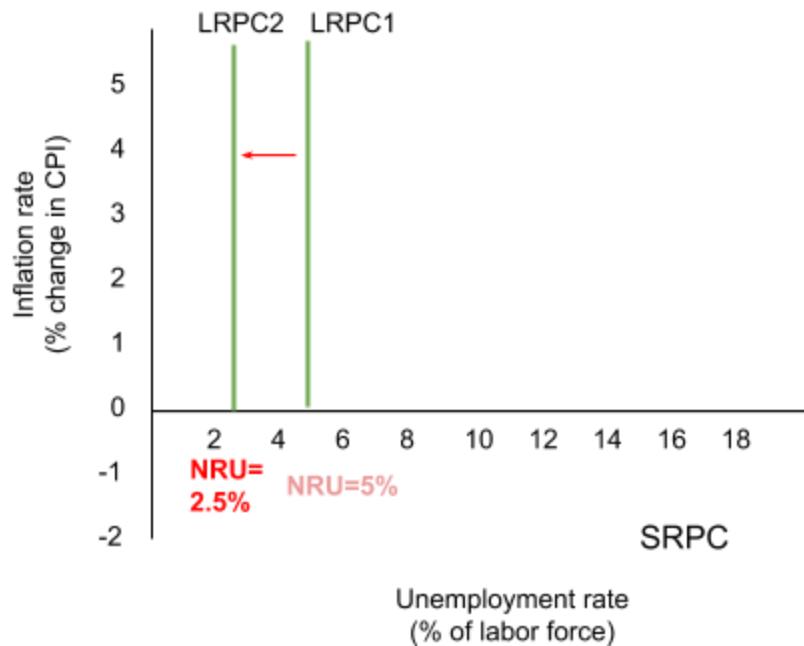
Factors that cause the natural rate of unemployment (NRU) to change will cause a shift of the LRPC. Recall that the NRU consists of two types of unemployment:

- **Structural unemployment** arises due to changing technology or other factors that result in a mismatch between the skills of a nation's workforce and the needs of employers.
- **Frictional unemployment** arises from workers who are in between jobs and cannot quickly and easily be matched up with firms that demand labor.

An increase in structural or frictional unemployment cause an outward shift of the LRPC and a higher natural rate of unemployment.



A decrease in structural or frictional unemployment causes an inward shift of the LRPC and a lower natural rate of unemployment.



Objective #3 - Economic growth

- Define economic growth as an increase in real GDP.
- Describe, using a production possibilities curve (PPC) diagram, economic growth as an increase in actual output caused by factors including a reduction in unemployment and increases in productive efficiency, leading to a movement of a point inside the PPC to a point closer to the PPC.

- Describe, using a PPC diagram, economic growth as an increase in production possibilities caused by factors including increases in the quantity and quality of resources, leading to outward PPC shifts.
- Describe, using an LRAS diagram, economic growth as an increase in potential output caused by factors including increases in the quantity and quality of resources, leading to a rightward shift of the LRAS curve.
- Explain the importance of investment for economic growth, referring to investment in physical capital, human capital and natural capital.
- Explain the importance of improved productivity for economic growth.

HL only objective:

- Calculate the rate of economic growth from a set of data.

Throughout the course we have defined and discussed the macroeconomic objective of economic growth. To review, **economic growth** is defined as the increase in the real output of an economy over time. The best measurement of economic growth is the growth rate in real GDP per capita over time.

Real GDP per capita is the total output of a country divided by the country's population.

$$\text{Real GDP per capita} = \frac{\text{Real GDP}}{\text{population}}$$

Some economists advocate the theory of an **aggregate production function**, which says that a country's rate of economic growth is a function of the allocation of the factors of production, including labor, capital, land, and technology.

According to the aggregate production function, real GDP is a function of the available inputs in a country, including:

- **Physical capital:** the machines and tools used by firms to produce goods and services
- **Labor:** the amount of work provided by humans in the production of goods and services
- **Human capital:** the level of skills and education of the country's workforce
- **Knowledge:** the availability and accessibility of scientific knowhow
- **Social infrastructure:** the regulatory, business, legal, and cultural environment
- **Natural resources:** the availability of raw materials (minerals, forest resources, energy resources, etc.)

Increases in any of the factors above will increase a country's real GDP; decreases in any of these will cause real GDP to fall.

For example, assume a country's government increases spending on public schools, and over time the quality of education and the skills with which students graduate improve. This increases the **human capital** in the country, and over time improves the **productivity** of the nation's workforce. Increased productivity will lead to economic growth at a higher income

and living standard for the nation's households.

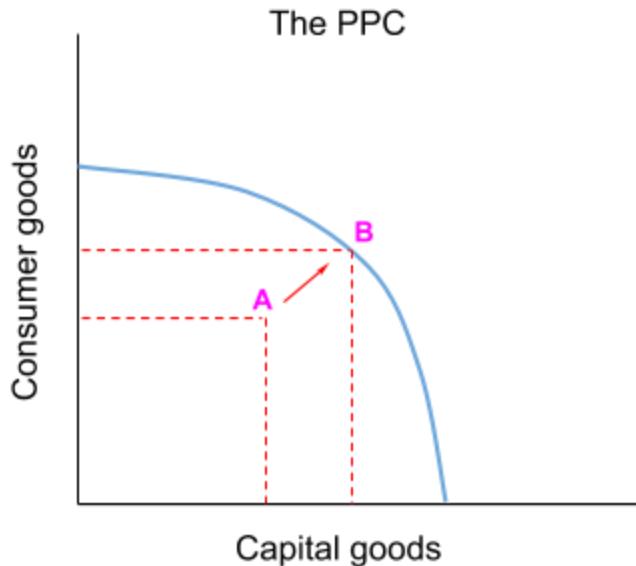
Other factors that can improve productivity and cause economic growth are

- technological improvements
- increases in physical capital

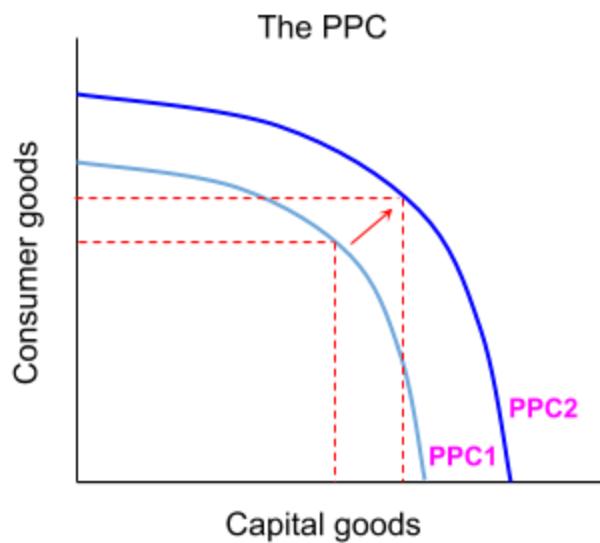
Sources of Long-run Economic Growth	
Physical Capital and Technology	<p>A nation's stock of physical capital is the quantity and quality of technology and infrastructure in a nation</p> <ul style="list-style-type: none"> ● Physical capital includes computer and communication technologies, high speed Internet access, efficient transportation networks, sanitation infrastructure and so on. ● Physical capital is only accumulated through INVESTMENT, either by the private sector or by the government. The private sector tends to provide ample investment in technologies that can easily be marketed and sold for a profit, while the public sector is needed to provide the public and merit goods that are necessary to make an economy grow: those things that would be under-provided by the free market, such as roads, ports, rail lines, communications infrastructure and so on.
Human Capital and Productivity	<p>Human capital refers to the quality of the labor force in a nation. Human capital can be improved through education and training of the nation's workforce, which may be undertaken by either the private sector or the government, or as is often the case, by both.</p> <ul style="list-style-type: none"> ● The quality of the education provided by a nation's school system determines the quality of the human capital produced, and therefore the productivity of workers. ● Greater skills among workers, and access to technology, increase the output per worker, and therefore the average income of households in a nation.

Illustrating economic growth

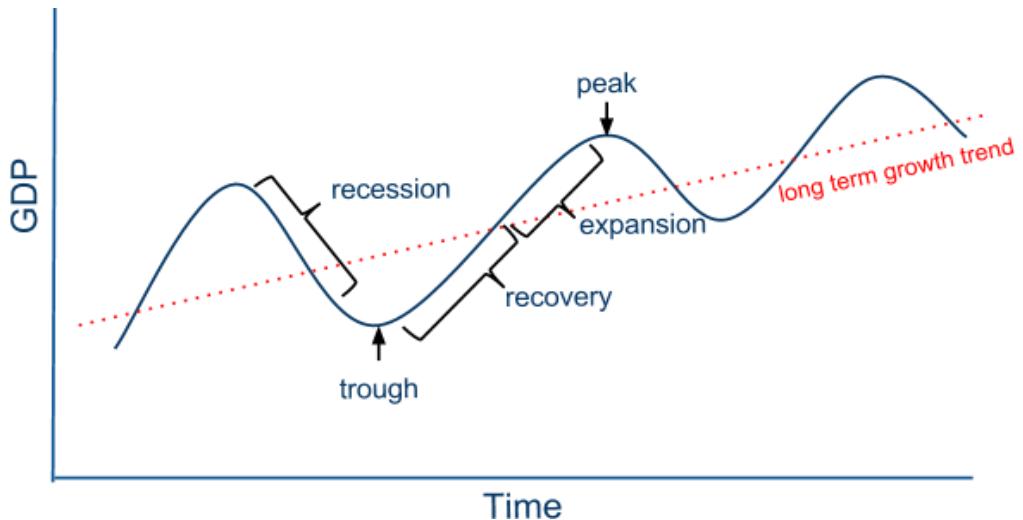
We have learned three models that show a country's long-run economic growth. Economic growth is shown as an increase in actual output caused by factors including a reduction in unemployment and increases in productive efficiency, leading to a movement of a point inside the **production possibilities curve (PPC)** to a point closer to the PPC, such as from A to B in the graph below.



Economic growth can also be illustrated as an increase in production possibilities caused by factors including increases in the quantity and quality of resources, leading to outward PPC shifts, as in the graph below.

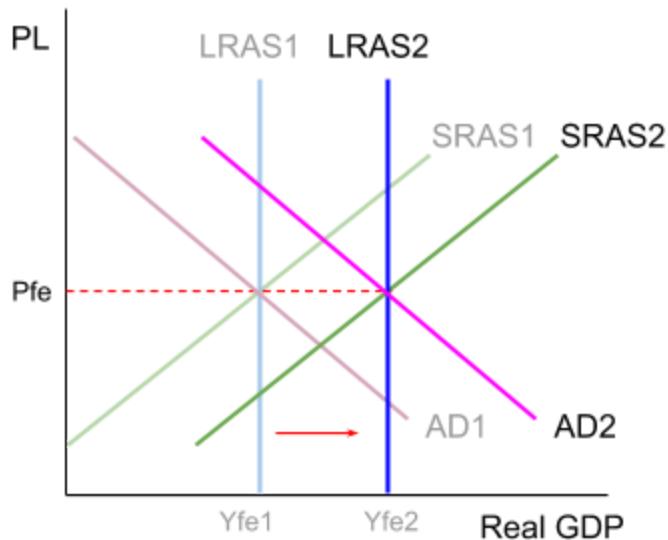


The **business cycle** shows economic growth as an increase in real GDP over time, shown by the “trend line” in the graph.



While an economy experiences fluctuations in short-run output, demonstrated by the phases of the business cycle, economic growth is shown by the red dashed line, which represents the increase in output experienced over the long run.

The **AD-AS model** shows economic growth as an outward shift of long-run aggregate supply and an increase in the full employment level of output.



An outward shift in the LRAS represents an increase in the full employment level of output in a nation. Furthermore, the increase in AD and SRAS show that total spending in the economy has increased with the level of output in the short run and the long run.

Calculating the rate of economic growth (HL only)

Economic growth can be represented mathematically as the percentage change in real GDP between two periods of time.

$$\text{Economic growth rate} = \frac{\text{real GDP in year 2} - \text{real GDP in year 1}}{\text{real GDP in year 1}} \times 100$$

The chart below shows real GDP in the United States from 1990 - 2016.



From the chart we can calculate economic growth in the US over the 26 year period:

- 2016 real GDP = \$17,000 billion
- 1990 real GDP = \$9,000 billion

$$\begin{aligned}\text{Economic growth from 1990-2016} &= \frac{2016 \text{ real GDP} - 1990 \text{ real GDP}}{1990 \text{ real GDP}} \times 100 \\ &= \frac{17,000 - 9,000}{9,000} \times 100 \\ &= \frac{8,000}{9,000} \times 100 \\ &= 88.9\%\end{aligned}$$

The US economy grew 88.9% between 1990 and 2016. The average rate of economic growth over the 26 year period can be calculated by dividing the total amount of growth by the number of years.

$$\text{Average growth rate} = \frac{88.9\%}{26} = 3.42\%$$

The US economy grew by an average of 3.42% per year from 1990-2016. In other words, the total output of goods and services produced by the country increased by 3.42% per year.

Consequences of economic growth

- Discuss the possible consequences of economic growth, including the possible impacts on living standards, unemployment, inflation, the distribution of income, the current account of the balance of payments, and sustainability.

Economic growth has some obvious consequences for the people of a country:

- **Higher incomes:** National output equals national income, so more output per person equates to higher average incomes.
- **Higher living standards:** More output per person generally means a higher living standards for the typical household. There are exceptions, however, if economic growth is accompanied by negative consequences such as environmental degradation or increasing inequality in society.
- **More jobs:** As an economy's output increases, the need for labor also increases, creating more employment opportunities for the country's households.

The most obvious effect of economic growth is higher incomes and greater output, both which benefit the households and firms in a nation. However, there are several undesirable consequences that may arise from the growth in output of a nation over time. These include:

- **Externalities:** Economic growth often comes at the expense of the environment. If growth is fueled by resource depletion, it may be unsustainable and may result in harmful effects on human and environmental health.
- **Inflation:** In economies experiencing rapid growth, high inflation often accompanies it. This means that if household incomes do not keep up with inflation, higher incomes may not actually improve standards of living over time. To adjust for inflation, it is important to consider the real economic growth per capita to determine how much better or worse off the typical household is from growth.
- **Structural unemployment:** A common effect of growth in the era of globalization is large numbers of people becoming structurally unemployed, as certain skills are no longer demanded in rapidly growing economies. An example is the disappearance of manufacturing jobs in the West as Asian nations attracted most new investment by manufacturers due to lower labor costs.
- **Composition of output:** If growth is primarily because of higher output in sectors that detract from human welfare (such as the weapons industry), then it may not make the typical household better off.
- **Unequal distribution of income:** the benefits of growth may not be shared across all segments of society, particularly if the rich see incomes rise dramatically while the middle class stagnates.
- **Trade balance:** In section 3 of the course you will learn about the **balance of payments**, which measures trade and investment flows between a country and the rest of the world. Economic growth can have negative impacts on a country's balance of trade, as rising incomes at home lead to more imports and less exports. The impact of growth on trade flows will be explored more in a later unit.

Objective #4 - Equity in the distribution of income

- Explain the difference between equity in the distribution of income and equality in the distribution of income.
- Explain that due to unequal ownership of factors of production, the market system may not result in an equitable distribution of income.

When thinking about **inequality**, it is important to distinguish between two concepts: equality and equity:

- **Equality** means that everyone earns the same amount, regardless of what skills he or she provides to the labor market. This was an objective of socialist economies based on the communist system of economics. This, however, is not an objective of a modern, market economy.
- **Equity** refers to fairness in the distribution of income. Increased fairness can be achieved through policies that, for example, place higher tax rates on higher income earners than on lower income earners, and through policies that redistribute the nation's income through transfer payments and spending on public goods in a way that provides equal opportunity for all members of society to earn a decent income, the government can promote greater equity.

The market system may not result in an equitable distribution of income. When resources are owned by private individuals and firms, it is inevitable that certain individuals will own more than others, thus the amount of income derived from land and capital (rent, interest, and profit) will be unevenly across society. Additionally, higher labor income (wages) are enjoyed by those who have been given access to a better education.

All this means that in free market economies, there will be a gap between high and low income earners. The lack of access to opportunity for economic advancement is a failure of the market system, since such advancement requires equitable access to education, health care, and other merit goods and services that tend to be underprovided by the free market.

Indicators of income equality/inequality

- Analyse data on relative income shares of given percentages of the population, including deciles and quintiles.
- Draw a Lorenz curve and explain its significance.
- Explain how the Gini coefficient is derived and interpreted.

To measure the income distribution of a nation, economists use a model known as the **Lorenz curve**, which shows the percentage of total income earned by the different segments of a nation's population.

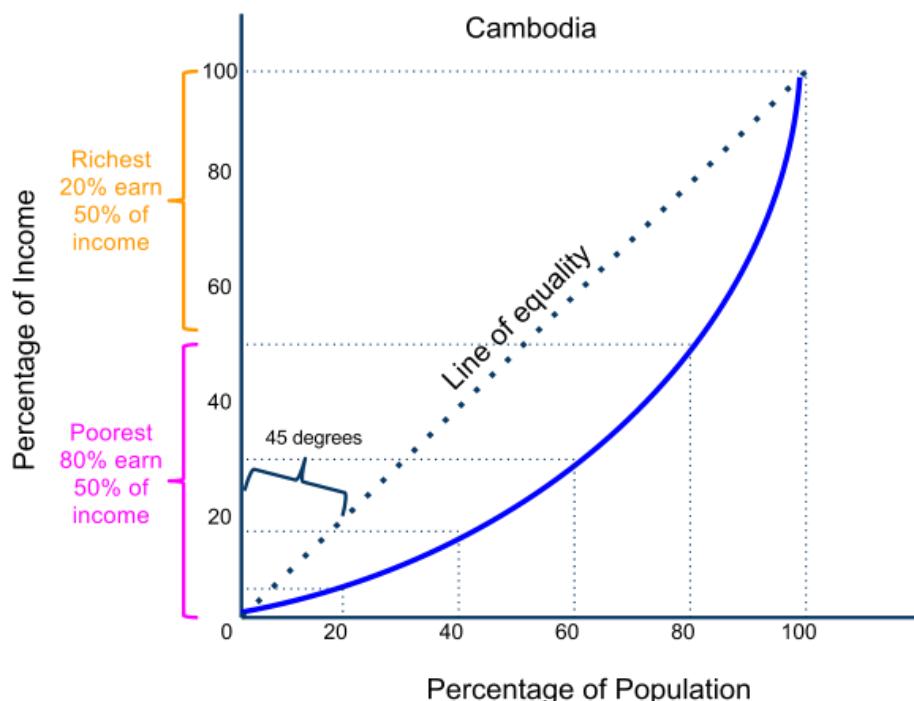
The Lorenz curve model plots the percentage of a nation's total income against the percentage of the nation's population, and thereby shows how much each quintile (or one fifth) of the population earns of the total income. Consider the table below:

% Of Population	Poorest 20%	2nd 20%	3rd 20%	4th 20%	Richest 20%	Gini Index
Cambodia	6.5%	9.7%	12.9%	18.9%	52%	43
Indonesia	7.4%	11%	14.9%	21.3%	45.5%	39.4
Brazil	3%	6.9%	11.8%	19.6%	58.7%	56.7
Vietnam	7.1%	10.8%	15.2%	21.6%	45.4%	37
Turkey	5.4%	10.3%	15.2%	22%	47.1%	41

The table shows us the percentages of total income earned by each quintile of the populations of several countries. From this data we can observe:

- Brazil is the least equal country on the list, because the poorest people earn a smaller proportion of total income relative to the richest compared to any other country.
- Indonesia and Vietnam are the most equal countries on the list. The poorest 20% earn a larger proportion relative to the richest 20% compared to the other countries.

The data from the table can be plotted in a Lorenz curve diagram. Let's take one country from the table, Cambodia, and plot the data on a graph with the percentage of the population on the horizontal axis and cumulative percentage of total income on the vertical axis. This is our Lorenz curve model.



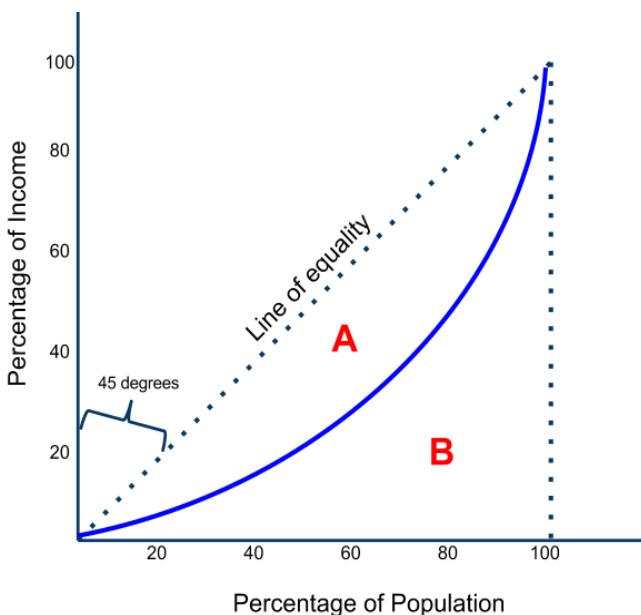
The curved line is Cambodia's Lorenz curve:

- It shows that the poorest 20% earn just 6.5% of total income, while the riches 20%

- earn over 50% of income.
- The 'line of equality' is for comparison; a country with a Lorenz curve along this line would be totally equal, with everyone earning the same amount.
- The further a country's Lorenz curve is from the line of equality, the more unequally income is distributed.
- The closer the Lorenz curve to the line of equality the more equally income is distributed.

The Gini coefficient as an indicator of income inequality

Because it would be inefficient to always draw a Lorenz Curve when comparing the levels of income equality across countries, economists have devised a numerical indicator of equality, known as the **Gini coefficient**.



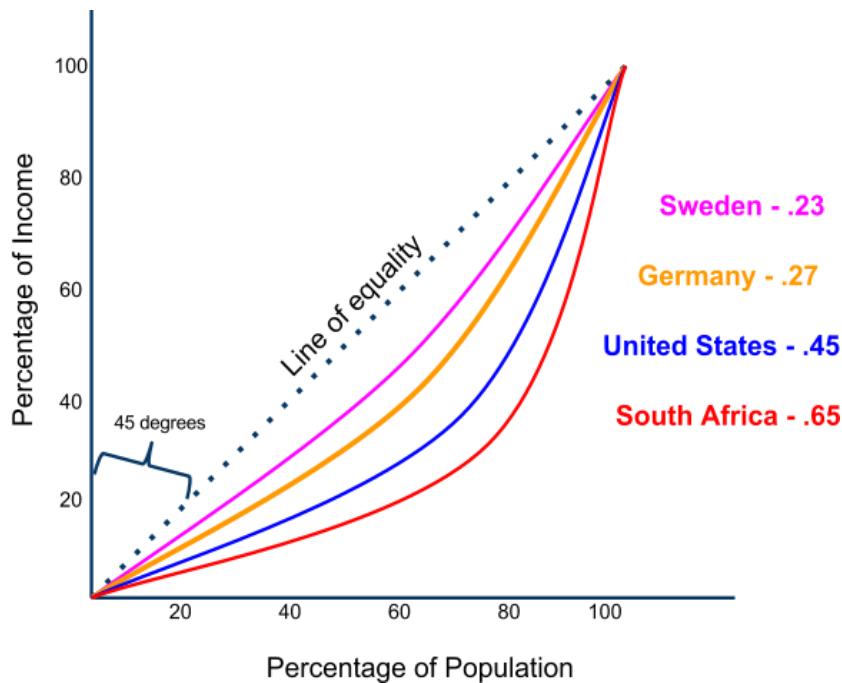
The Gini coefficient is the ratio of the area below the line of equality and the Lorenz curve and the entire area below the line of quality.

$$\text{Gini coefficient} = \frac{\text{area A}}{\text{area A+B}}$$

- Notice that the further away the Lorenz curve is from the line of equality, the greater A will be relative to B, and therefore the higher the Gini coefficient will be.
- The closer the curve to the line of equality, the smaller A will be relative to B, and the lower the Gini coefficient will be.
- The closer the coefficient is to ZERO, the more equal a country is. The closer it is to ONE, the less equal the country is

Comparing income distributions using Lorenz curves

By comparing the Lorenz curves and Gini coefficients of different countries, we can draw conclusions about the levels of income inequality.



From the Diagram, we can conclude:

- Sweden is the country in which income is distributed most evenly. Its Gini coefficient is smaller, indicating that its Lorenz curve is closer to the line of equality.
- South Africa is the country in which income is distributed least evenly. Its Gini coefficient is higher, indicating its Lorenz curve is further from the line of equality.

Sources of income inequality

Income inequality has several causes, some of which are summarized here:

- Unequal distribution of resources: If a large proportion of a country's factors of production (capital and land) are owned by a relatively small proportion of citizens, then the majority of income will be earned by those individuals, while the majority of society will share a small chunk of total income.
- Inequitable access to education: If a country's education system fails to provide everyone in society, regardless of social standing or income level, with a quality education, then the quality of human capital will be unequal across society, resulting in lower incomes among those for whom quality education is unavailable.
- Underprovision of merit goods like healthcare: If certain segments of the population are less healthy than others and do not have access to healthcare, then they will be less productive and earn less income than others in society. The provision of healthcare by government can eliminate the inequities arising from the underprovision of merit goods.
- Regressive tax system: A country that relies primarily on regressive taxes such as indirect sales taxes for its government revenue will likely have greater income inequality than a country that depends more on progressive taxes such as an income tax. The distinction between different tax systems is explored later in this chapter.

There are many other sources of income inequality that may explain the large gap between the rich and poor in society. Historical factors such as the legacy of colonialism and cultural factors such as the racial and ethnic inequalities characteristics in many countries can also explain the gaps between rich and poor.

Poverty

- Distinguish between absolute poverty and relative poverty.
- Explain possible causes of poverty, including low incomes, unemployment and lack of human capital.
- Explain possible consequences of poverty, including low living standards, and lack of access to health care and education.

Without certain government policies to promote equity, a free market economy will ultimately create a society in which there is a vast gap between the richest and the poorest households. This will likely result in poverty - both relative poverty and, in the poorest countries, absolute poverty.

Relative poverty exists whenever certain households in a nation earn an income that makes them poor relative to the richer households in a nation. Relative poverty exists in even the richest countries, and while individuals may be able to afford the basic necessities of life, their standards of living will be significantly lower than the relatively rich within their societies.

Absolute poverty exists when a household earns an income that is below a level that allows them to buy even the basic necessities of life (nutritious food, shelter, clothing, education and health).

- The World Bank defines the absolute poverty line as \$2 per day. There are approximately 1 billion individuals in the world living in absolute poverty in 2012, mostly in Sub-Saharan Africa, parts of Central Asia and parts of Latin America.
- Unlike relative poverty, which exists everywhere, there is little or no absolute poverty in the more developed countries in the world.

The existence of both relative and absolute poverty in a country has consequences for the economy and society as a whole. Individuals in poverty...

- have low living standards
- are unable to access education and healthcare
- are less productive, thus contribute less to the country's economic activity
- may place a burden on others in society by depending on social safety nets (unemployment and welfare benefits)

Reducing poverty is a worthwhile macroeconomic objective for the social, economic, and humanitarian benefits such reductions will create.

The role of taxation in promoting equity

- Distinguish between direct and indirect taxes, providing examples of each, and

- explain that direct taxes may be used as a mechanism to redistribute income.
- Distinguish between progressive, regressive and proportional taxation, providing examples of each.

HL only objective:

- Calculate the marginal rate of tax and the average rate of tax from a set of data.

Certain government interventions in the free market can promote equity, or fairness, in a market economy, and thus reduce the level of income inequality in a nation. Policies include the **taxes** a government collects and the **transfer payments** and **public goods** it provides.

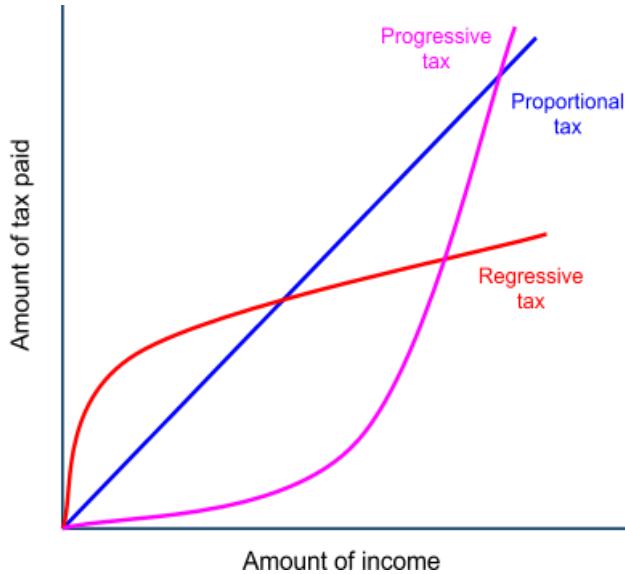
A tax can be either regressive, proportional or progressive:	
Regressive taxes	A tax that consist of a larger percentage of poor household's income than a rich household's income. <ul style="list-style-type: none"> Indirect taxes are regressive: A tax on consumption of a particular percentage places a larger burden on a poor consumer than on a rich consumer. For example, a 10% tax on a \$100 good is \$10. To a poor household, \$10 is a much bigger deal than to a rich household. A government that uses indirect taxes as its primary revenue source will actually worsen income inequality in society since they place a smaller burden on the rich than on the poor.
Proportional taxes	This is a tax that remains constant as a proportion of income as incomes increase. A direct tax (on income) may be a flat tax, for example, all households pay 15% of their income in tax, regardless of the income level. Proportional taxes do nothing to promote income equality, although they will not worsen inequality
Progressive taxes	A tax that increases as a percentage of total income as income increases. For example, households earning \$50,000 may pay only 10% in income tax, while those earning \$250,000 may pay 35% in taxes. A progressive tax promotes greater income equality because those who can afford to pay the most do, and those whose incomes are lower and cannot afford to pay as much pay less, leaving poorer households with more disposable income to enable a higher standard of living.

The role of taxes in affecting income distribution

A **progressive income tax** is the most proven method for reducing income inequality.

It is not a coincidence that the most equal countries in the world (Sweden, Denmark, Norway, France, the Netherlands) have some of the most progressive tax systems.

The graph below shows the amount of tax paid relative to income under the three systems:



Observe the following from the graph above:

- Under **regressive taxes** the rich pay MORE total tax (since they consume more than the poor), but as a proportion of their income they pay LESS as their incomes rise (since richer households tend to save a larger proportion of their incomes than poor households)
- Under **proportional taxes** (sometimes called ‘fair taxes’ because everyone pays the same percentage), the amount of tax increases proportionally with income. Everyone pays the same percentage. Under such a system the taxes do not increase equality or provide greater equity. Income is not redistributed since everyone faces the same tax burden, despite their level of income.
- **Progressive taxes** increase exponentially as income increases. Arguments against the progressive tax say it discourages hard work, because the more income you earn, the more you have to pay in tax. A progressive tax system does the most to promote equity and equality in income distribution, since the poor face the smallest burden and the rich face the largest burden.

Calculating marginal and average tax rates (HL only)

A progressive income tax, common in most countries, increases in percentage as income rises. For example, the table below shows the “income tax brackets” for a hypothetical country.

Income (\$ per year)	Marginal rate of taxation
1-10,000	0%
10,001-30,000	15%
30,001-50,000	25%

50,001-80,000	35%
80,001 and above	50%

From the table we can distinguish between the marginal tax rate and the average tax rate paid by four hypothetical individuals in the country.

- Individual A earns \$15,000
- Individual B earns \$36,000
- Individual C earns \$64,000
- Individual D earns \$200,000

Below we will calculate how much tax each of the individuals will pay and the average tax rates paid by each individual.

Individual A: \$15,000

- The first \$10,000 are tax free
- On the next \$5,000 the individual pays 15%
- Total tax = $\$5,000 \times 0.15 = \750

Individual A pays \$750 in taxes. Her average tax rate can be calculated by dividing her total tax by her income and multiplying the result by 100.

$$\text{Average tax rate} = \frac{\text{tax}}{\text{income}} \times 100$$

Individual A's average tax rate is:

$$\frac{\$750}{\$15,000} \times 100 = 5\%$$

We can calculate the amount of tax and average tax rates paid by the other individuals:

Individual B: \$36,000

- First \$10,000 are tax free
- $\$10,001 - \$30,000$ taxed at 15%: $\$20,000 \times 0.15 = \$3,000$
- $\$30,001-36,000$ taxed at 25%: $\$6,000 \times 0.25 = \$1,500$
- Total tax = $\$3,000 + \$1,500 = \$4,500$
- Average tax rate = $\frac{\$4,500}{\$36,000} \times 100 = 12.5\%$

Individual C: \$64,000

- First \$10,000 are tax free
- $\$10,001 - \$30,000$ taxed at 15%: $\$3,000$
- $\$30,001 - \$50,000$ taxed at 25%: $\$20,000 \times 0.25 = \$5,000$
- $\$50,001 - \$64,000$ taxe at 35%: $\$16,000 \times 0.35 = \$5,600$
- Total tax = $\$3,000 + \$5,000 + \$5,600 = \$13,600$

- Average tax rate = $\frac{\$13,600}{\$64,000} \times 100 = 21.25\%$

Individual D: \$200,000

- First \$10,000 are tax free
- \$10,001 - \$30,000 taxed at 15%: \$3,000
- \$30,001 - \$50,000 taxed at 25%: \$5,000
- \$50,001-\$80,000 taxed at 35%: $\$30,000 \times 0.35 = \$10,500$
- \$80,001-\$200,000 taxed at 50%: $\$120,000 \times 0.50 = \$60,000$
- Total tax = $\$3,000 + \$5,000 + \$10,500 + \$60,000 = \$78,500$
- Average tax rate = $\frac{\$78,500}{\$200,000} \times 100 = 39.25\%$

The marginal and average tax rates paid by the four individuals can be summarized in the table below:

Income	Marginal tax rate	Average tax rate
\$15,000	15%	5%
\$36,000	25%	12.5%
\$64,000	35%	21.25%
\$200,000	50%	39.25%

Notice that the average tax rate is always lower than the marginal rate in a progressive income tax system. This is due to the fact that the marginal rates only apply to different “chunks” of an individual’s income. The 50% rate, for example, only applies to the last \$120,000 of individuals D’s income. All the income below \$80,000 is taxed at a lower rate.

Other measures to promote equity

- Explain that governments undertake expenditures to provide directly, or to subsidize, a variety of socially desirable goods and services (including health care services, education, and infrastructure that includes sanitation and clean water supplies), thereby making them available to those on low incomes.
- Explain the term transfer payments, and provide examples, including old age pensions, unemployment benefits and child allowances.

The method a government uses to collect taxes (direct or indirect, regressive or progressive) only provides us with half the explanation of how government can promote equity. The other half of the equation is how the government spends its money.

To reduce inequality, government must allocate the money collected through progressive taxes in ways that increase equity in society. The provision of **public goods** and **merit**

goods, such as healthcare services, education, and infrastructure (including sanitation and clean water), assures that low income households are able to access these essential goods. Government provision levels the playing field so that goods essential to developing human capital are available to everyone in society, not just the richest individuals.

Another type of government intervention in the market to reduce inequality is transfer payments. A **transfer payment** is a payment from the government to an individual for which no good or service is exchanged, rather income is redistributed from one group to another. The government's provision of transfer payments is intended primarily to provide greater equity through increased opportunities for low income households.

Examples of equity-enhancing transfer payments include:

- Unemployment benefits: To limit the fall in income experienced by individuals who lose their jobs
- Social Security: To provide income to individuals in old age; reduces poverty among elderly
- Nutritional subsidies: To help low income households afford healthier food and improve health
- Educational subsidies: To help low income households afford a college education
- Health care subsidies: To reduce the cost of healthcare for lower income individuals
- Welfare benefits: To provide the poorest of households with an income so they can purchase the basic necessities

The relationship between equity and efficiency

- Evaluate government policies to promote equity (taxation, government expenditure and transfer payments) in terms of their potential positive or negative effects on efficiency in the allocation of resources.

Opponents of government interventions aimed at reducing income inequality, such as progressive taxes and transfer payments, argue that such measures reduce overall efficiency in the economy. The argument goes as follows:

- Progressive income taxes, which increase in percentage as incomes rise, create a disincentive for individuals to pursue more productive, higher paying jobs, as doing so will mean they pay a larger share of their income in taxes. As a result, workers will voluntarily limit their efforts and output in the workplace, reducing the rate of economic efficiency across the economy.
- Transfer payments such as unemployment and welfare benefits create a disincentive for people collecting them to pursue productive employment, since doing so may mean they cease to enjoy government benefits and may experience a fall in their incomes. Abolishing such transfer payments, the argument goes, would motivate more people to accept jobs at any wage rate and contribute productively to the economy.

The tradeoff between equity and efficiency is at the heart of the debate over the extent to which government should intervene in the free market to reduce inequality.

2.4 Fiscal Policy

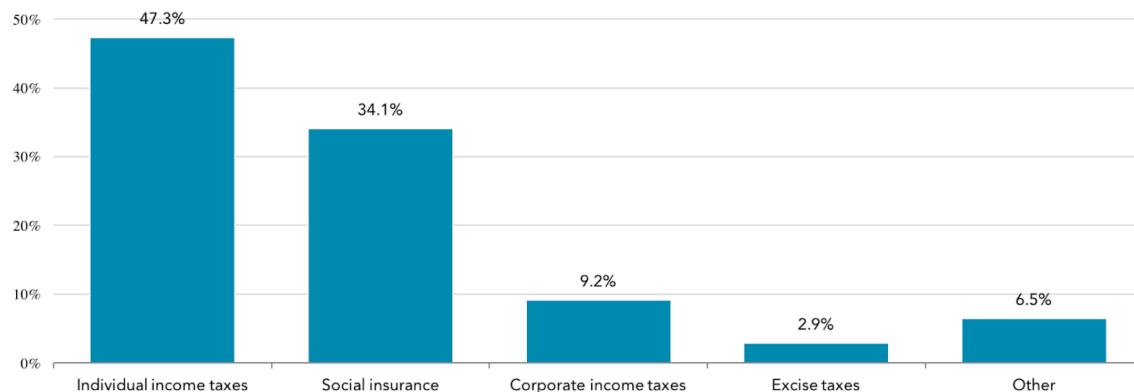
The government budget - sources of government revenue

- Explain that the government earns revenue primarily from taxes (direct and indirect), as well as from the sale of goods and services and the sale of state-owned (government-owned) enterprises.

A **government budget** consists of the value of revenues earned and expenditures made by the government in a period of time, usually a year. Sources of a government's revenues include:

- Taxation:** Direct (income) taxes and indirect (consumption) taxes make up the vast majority of most governments' revenues. The main sources of tax revenue include
 - Personal income taxes - taxes on individual employees' incomes
 - Corporate taxes - taxes on business firms' incomes
 - Payroll taxes - proportional taxes shared by employers and employees meant to cover pensions, unemployment benefits, disability benefits, and other transfer payments
 - Excise taxes - taxes on particular goods, such as automobile fuel, tobacco, alcohol, including taxes on imported goods.
 - Value-added (sales) taxes - a tax places on consumption of all goods. Note that some countries do not have national sales taxes, only local.
- Sale of state-owned enterprises:** When a government privatizes state-owned assets such as land or capital, it earns revenues from the sale of these assets to private investors or firms.

The table below, from the Tax Policy Center, shows the sources of US government revenues for 2016. (note: "social insurance" refers to payroll taxes).



Taxes account for all but 6.5% of the revenue earned by the US government in 2016.

Types of government expenditures

- Explain that government spending can be classified into current expenditures, capital expenditures and transfer payments, providing examples of each.

Once a government has collected revenues, it in turn spends money on public goods and services and transfer payments. **Government expenditures** are classified as either current or capital expenditures or transfer payments.

- Current expenditures** refer to government spending on wages for government employees (including military personnel) and raw materials. Current expenditures must be renewed each year just to keep the government functioning.
- Capital expenditures** refer to government spending on fixed capital such as roads, bridges, hospital, schools, government building, and other infrastructure. Capital expenditures are long-term and do not need to be renewed each year.
- Transfers** are payments from the government to individuals or businesses for which no goods or services are provided. For example:
 - Unemployment benefits
 - Social security (pension) payments
 - Producer subsidies
 - Welfare payments

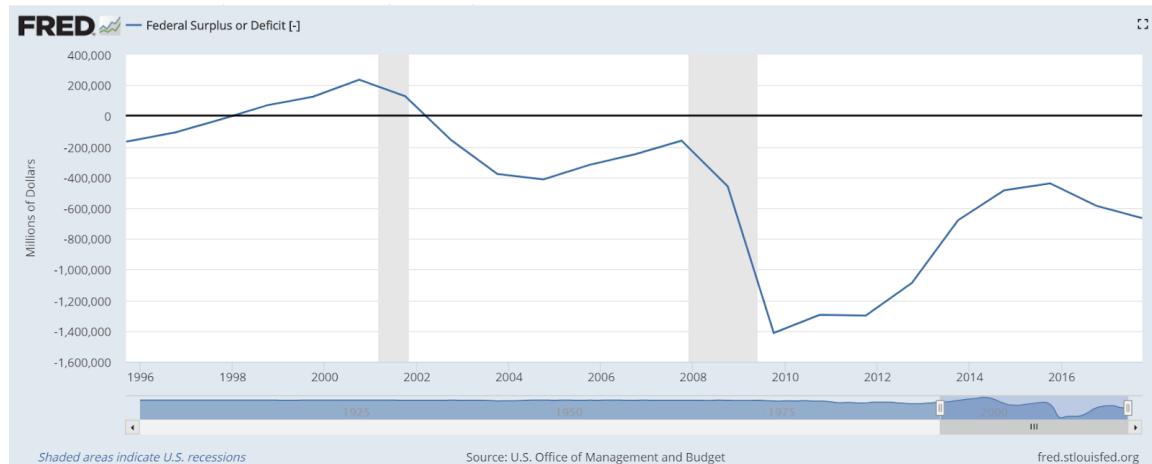
The budget outcome

- Distinguish between a budget deficit, a budget surplus and a balanced budget.
- Explain the relationship between budget deficits/ surpluses and the public (government) debt.

A government's budget can be in either surplus or deficit.

- A **budget surplus** is when a government's total receipts from tax revenues are greater than its total spending on public goods and transfers.
- A **budget deficit** is when a government's total spending on public goods and transfers is greater than total receipts from tax revenues.

The government budget can be in surplus or deficit in a given time period. The chart below shows the US government budget balance over 20 years, from 1996 to 2016.



The black line represents a **balanced budget**, when tax receipts equal government expenditures. When the blue line is above the balanced budget line, the budget is in surplus, and when it's below the balanced budget line the budget is in deficit. Gray areas indicate years in which the US economy was in recession. Observe the following from the chart:

- The US government budget was in deficit in 1996-1997, then every year since 2003.
- The budget was balanced in 1998 and 2002.
- The budget was in surplus in 1999, 2000, and 2001.

During the 2008-2009 recession (the “Great Recession”), the budget deficit ballooned from \$200 billion to \$1,400 billion (or \$1.4 trillion).

The relationship between the budget balance and fiscal policy

A government’s fiscal policy has direct and immediate impacts on its budget balance.

- **Expansionary fiscal policies**, when taxes are cut or government spending is increased, move the budget towards deficit. If the budget is in surplus, the surplus will shrink or move to deficit. If the budget is already in deficit, the size of the deficit will increase.
- **Contractionary fiscal policies**, when taxes are increased or spending is cut, move the budget towards surplus. If the budget is in deficit, the deficit will shrink as tax receipts increase and spending is cut; if the budget is already in surplus, the size of the surplus will grow.

The relationship between the budget balance and the national debt

A country’s **national debt** is the sum of all past budget deficits minus surpluses of a government over time. It is how much the government owes the public, including domestic and international lenders (bondholders), from all the borrowing the government has done to finance past budget deficits.

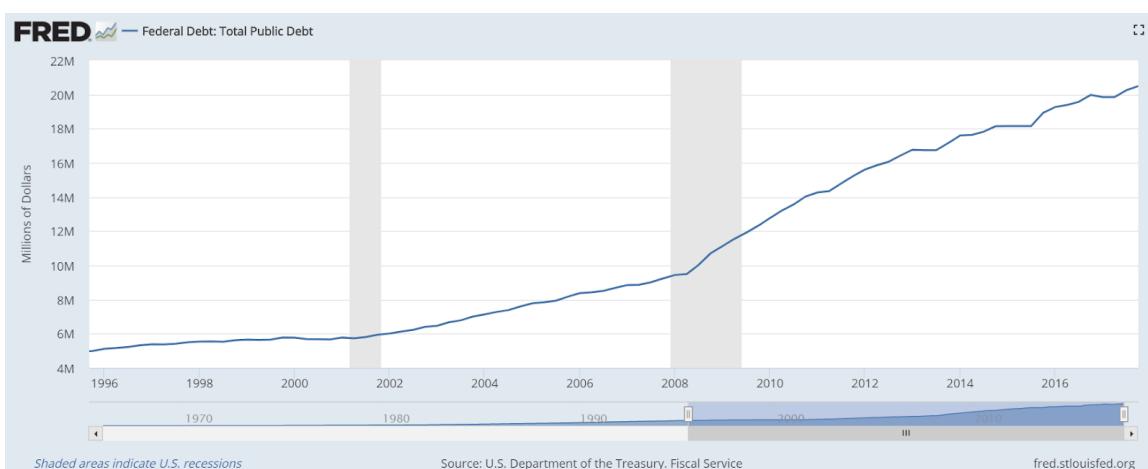
When the government’s budget is in deficit, it must borrow funds to **finance the deficit**. It is the same as when an individual spends more than his or her income. Imagine you get \$50 per week in allowance from your parents, but this week your allowance runs out on

Thursday. In order to finance continued spending, you must borrow additional funds from your parents, so you borrow an additional \$20 for the weekend. Your week's deficit is now \$20, and this increases whatever past debt you owe your parents by \$20.

Perhaps next week you will only spend \$30 of your allowance, freeing up \$20 to repay your debts. In week 2 you had a budget surplus of \$20, which allowed you to decrease your debt by paying your parents back the \$20 you owed them.

In years a government has a budget deficit, its total debt increases. When the budget is in surplus, the government has the opportunity to reduce its debt by paying back past lenders.

The graph below shows the US government's total debt from 1996 to 2016.



Notice that in the years in which the budget was balanced or in surplus (1998-2002), the total public debt remained constant at about \$6 trillion. However, from 2003 onwards, when the budget was in deficit every year, total debt grew steadily to over \$20 trillion in 2016.

Also notice that during the Great Recession (2008-2009), the rate of increase in total debt was greatest. This is due to the fact that during these years the budget experienced its largest deficits in history (over \$1 trillion). The US government enacted an expansionary fiscal policy to combat the Great Recession, resulting in a fall in tax receipts and a simultaneous increase in government spending.

Fiscal policy and short-term demand management

- Explain how changes in the level of government expenditure and/or taxes can influence the level of aggregate demand in an economy.
- Describe the mechanism through which expansionary fiscal policy can help an economy close a deflationary (recessionary) gap.
- Construct a diagram to show the potential effects of expansionary fiscal policy, outlining the importance of the shape of the aggregate supply curve.
- Describe the mechanism through which contractionary fiscal policy can help an

- economy close an inflationary gap.
- Construct a diagram to show the potential effects of contractionary fiscal policy, outlining the importance of the shape of the aggregate supply curve.

Changes in the level of taxation, transfers, and government spending will impact the level of aggregate demand in the economy and affect the short-run equilibrium level of output.

Changes in government spending impact AD directly, while changes in taxes and transfers affect AD indirectly. Consider, for example the following fiscal policies:

- **Government increases spending on new infrastructure by \$10 billion:** Increased infrastructure spending creates a direct stimulus to AD in the economy. Incomes will increase by \$10 billion, and then further increases in spending will multiply the initial change, resulting in an ultimate increase in GDP of more than \$10 billion.
- **Government cuts household income taxes by \$10 billion:** A tax cut will have an indirect impact on AD. The initial \$10 billion tax cut will increase households' disposable incomes, but only some proportion of that will turn into new spending in the economy. For example, if the marginal propensity to consume is 0.8, then a \$10 billion tax cut will cause an initial change in spending of $\$10 \text{ billion} \times 0.8 = \8 billion . The \$8 billion of new spending will then be multiplied throughout the economy until the ultimate change in GDP is greater than the initial tax cut.
- **Government increase spending on unemployment benefits by \$10 billion:** An increase in transfers, such as unemployment benefits, will also affect AD indirectly. The \$10 billion increases disposable incomes among the country's unemployed individuals, who will then spend some proportion of that.

Changes in taxation and transfers have an indirect effect on AD, since they depend on households to determine how much of the tax cut or increase in transfers will turn into new spending. A change in government spending, on the other hand, impacts AD directly, since government spending (G) is one of the components of AD.

Fiscal policy and the spending multiplier

The impact of a particular change in government spending will always be greater than a change in taxation or transfers of the same amount. As explained above, government spending is a direct component of AD, while taxation and transfers only affect AD indirectly. To better understand this, we can revisit the spending multiplier introduced earlier in the course.

For example, assume the marginal propensity to consume (MPC) in a country is 0.6 . With this values we can calculate the spending multipliers:

$$\text{The spending multiplier} = \frac{1}{1-\text{MPC}} = \frac{1}{0.4} = 2.5$$

With this value we can calculate the impact of a change in government spending or taxation. For example, assume government reduces spending on defense by \$10 billion. The impact on overall AD can be calculated:

$$\begin{aligned}
 \text{Total } \Delta \text{ in AD} &= \text{initial } \Delta \text{ in spending} \times \text{spending multiplier} \\
 &= -\$10 \text{ billion} \times 2.5 \\
 &= \textcolor{red}{-\$25 \text{ billion}}
 \end{aligned}$$

A \$10 billion decrease in government spending will result in a \$25 billion decrease in AD.

Next let's consider the impact of a \$10 billion increase in taxes.

$$\begin{aligned}
 \text{Total } \Delta \text{ in AD} &= (\text{change in taxes} \times -\text{MPC}) \times \text{spending multiplier} \\
 &= (\$10 \text{ billion} \times -0.6) \times 2.5 \\
 &= \textcolor{red}{-\$15 \text{ billion}}
 \end{aligned}$$

The change in taxes must first be multiplied by the MPC, since only that proportion of the tax change will impact spending in the economy (since a change in taxes also affects imports, taxes, and savings).

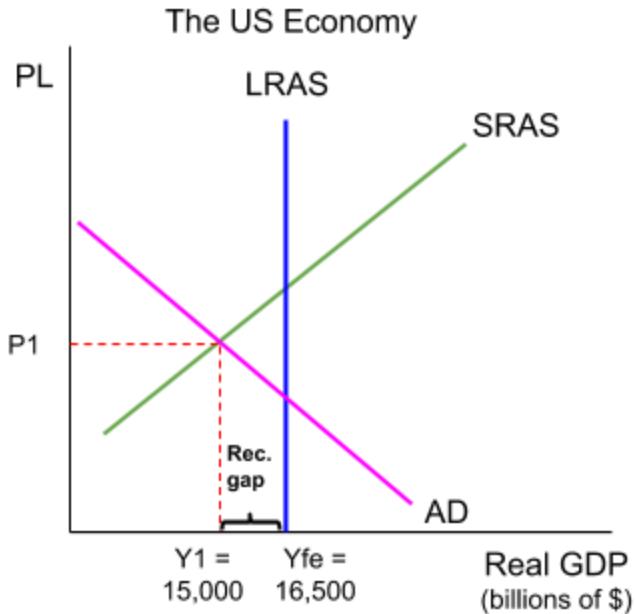
Notice that a \$10 billion dollar decrease in government spending has a larger impact on total spending than a \$10 billion increase in taxes. Again, this reflect that a change in government spending affects AD directly while a change in taxes affects AD indirectly.

Of course, this holds true when fiscal policy is used to close negative output gaps as well. For example, if there is a \$25 billion negative output gap, then a \$10 billion increase in spending should be enough to restore full employment (assuming an MPC of 0.6), but a \$10 billion decrease in taxes will not.

Expansionary fiscal policies

During a recession, a government can implement an **expansionary fiscal policy**. Either a **tax cut** or an **increasing in government spending** will stimulate aggregate demand and help close a recessionary gap.

Consider the economy below.

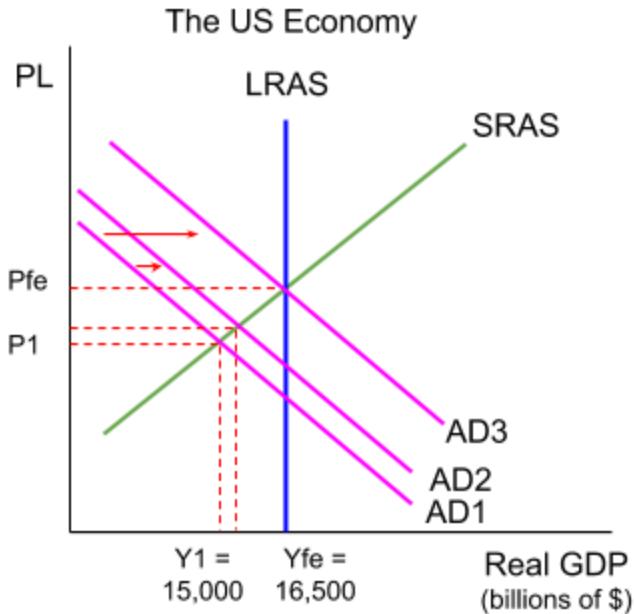


The economy is producing \$1.5 trillion below its full employment level of output. To stimulate demand, an expansionary fiscal policy could be employed. Assume the MPC = 0.6.

To determine the appropriate size of an increase in spending, we can divide the desired change in GDP by the spending multiplier.

$$\begin{aligned}
 \text{Needed change in spending} &= \frac{\text{desired } \Delta \text{ in GDP}}{\text{spending multiplier}} \\
 &= \frac{\$1.5 \text{ trillion}}{2.5} \\
 &= \textbf{\$0.6 trillion (\$600 billion)}
 \end{aligned}$$

With a \$1.5 trillion recessionary gap and a spending multiplier of 2.5, government can expect a \$600 billion increase in spending to stimulate AD to a level that would result in the economy returning to full employment. The impact of the fiscal stimulus can be shown in the AD-AS model.



The \$600 billion increase in government spending initially shifts AD from AD1 to AD2. As incomes increase because of the new government spending, consumer spending increases and the initial change in spending is multiplied 2.5 times throughout the economy. Ultimately, AD shifts out to AD3, restoring full employment output at Y_{fe} .

A tax cut could also have been used to stimulate AD and help restore full employment. To determine the size of the tax cut required, we can divide the desired change in GDP by the spending multiplier, divide that by the multiplier again.

$$\begin{aligned} \text{Needed change in taxes} &= \frac{\text{desired } \Delta \text{ in GDP}}{\text{spending multiplier}} \div -\text{MP C} \\ &= \frac{\$1.5 \text{ trillion}}{2.5} \div -0.6 \\ &= \textbf{-\$1 trillion} \end{aligned}$$

A \$1.5 trillion recessionary gap could realistically be closed with a tax cut of \$1 trillion. Such a tax cut will increase households' disposable incomes and result in an initial increase in spending of $\$1 \text{ trillion} \times 0.6$, or \$600 billion. This \$600 billion increase in spending is then multiplied by the spending multiplier (2.5) to achieve the desired \$1.5 trillion increase in AD.

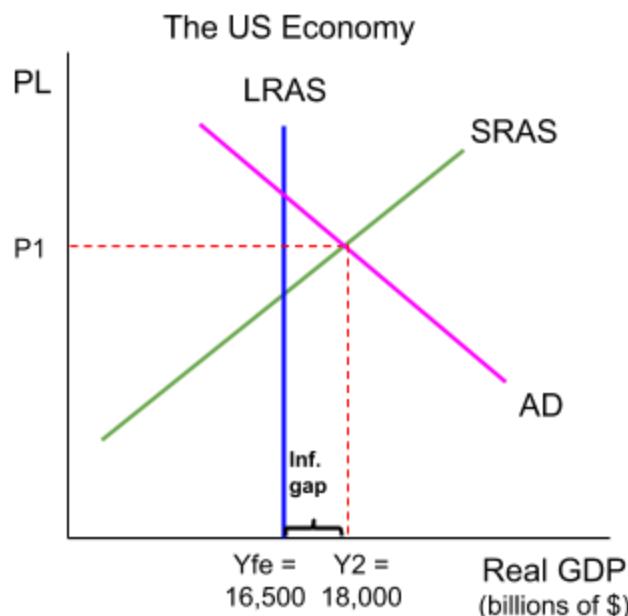
As we can see, in order to achieve an initial change in spending large enough to stimulate AD to its full employment level, a much larger tax cut is needed than the size of a government spending increase that would achieve the same result. Policymakers concerned about the impact of expansionary fiscal policies on the **government budget** should be aware of this difference. If the goal is to get the most "bang for their buck," government should favor spending increases over tax cuts.

However, in reality, policymakers often favor tax cuts over spending increases when implementing fiscal policy, likely because tax cuts are extremely popular with voters (many of whom would benefit from a better understanding of fiscal policy and multiplier theories!)

Contractionary fiscal policies

Recessions are not the only macroeconomic problem a country can face. If an economy is experiencing a positive output gap, then inflation is higher than desired and unemployment is below its natural rate. Such an economy is overheating, and a **contractionary fiscal policy** could be employed to reduce AD, reduce inflation, and restore full employment.

Consider the economy below.

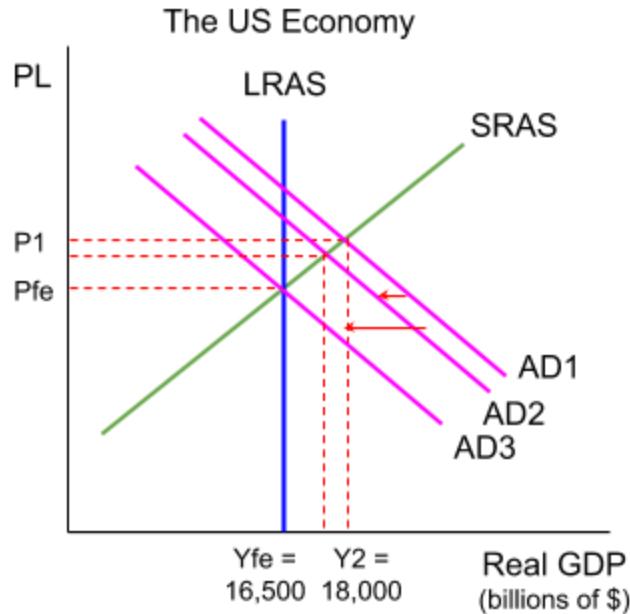


This economy is producing a level of output \$1.5 trillion above its full employment level. The unemployment rate is lower than the NRU and inflation is higher than desired. Either a decrease in government spending or an increase in taxes can be implemented to reduce AD and restore full employment.

A **decrease in spending** of **\$600 billion** will be multiplied by the spending multiplier of 2.5 and cause total spending in the economy to fall by \$1.5 trillion.

An **increase in taxes** of **\$1 trillion** will reduce households' disposable income and cause consumption to fall by \$600 billion, ultimately resulting in a \$1.5 trillion decrease in AD, restoring full employment.

The effects of a contractionary fiscal policy can also be illustrated in the AD-AS model.



The initial \$600 billion decrease in government spending (or \$1 trillion increase in taxes) causes AD to shift from AD1 to AD2. As incomes decrease and consumer spending decreases, the initial change in spending is multiplied throughout the economy until the ultimate impact on AD is a fall to AD3. Output returns to full employment at \$16.5 trillion and the price level returns to the full employment level of P_{fe} (consistent with a low and stable inflation rate).

The table below shows the changes in total spending that will result from an initial change in expenditures of \$100 in an economy based on different marginal propensities to consume.

Marginal Propensity to Consume	Size of the spending multiplier	Change in government spending	Total change in spending in the economy
0.2	1.25	+100	125
0.4	1.67	+100	167
0.6	2.5	+100	250
0.8	5	+100	500
0.9	10	+100	1000

The impact of automatic stabilizers

- Explain how factors including the progressive tax system and unemployment benefits, which are influenced by the level of economic activity and national income, automatically help stabilize short-term fluctuations.

When government policymakers get together and decide to take fiscal actions to either

contract or expand aggregate demand, they are implementing **discretionary fiscal policies**. However, when changes in output or employment trigger automatic increases or decreases in spending and taxation, an economy is employing **automatic stabilizers**.

Automatic stabilizers are the built-in changes in transfers and taxation that happen when an economy's output and employment increase or decrease.

Automatic stabilizers during a recession

Consider what happens following a fall in AD resulting from a decrease in net exports:

- As AD falls, workers in the export sector are laid off due to sticky wages and other input costs in the short run.
- Unemployed workers who have lost their income stop paying income taxes, so there is an automatic decrease in the amount of tax collected by the government.
- As long as AD remains weak, workers will remain unemployed and many will begin collecting unemployment benefits from the government; therefore, there is an automatic increase in transfers as AD falls.

During a demand-deficient recession tax revenues decrease automatically as GDP falls, preventing consumption and the economy from falling further. At the same time, transfer payments increase, helping maintain consumption among those who have lost their jobs. The effect of these automatic stabilizers is to, well, stabilize aggregate demand and prevent it from falling as much as it would if these effects did not kick in.

Automatic stabilizers during demand-pull inflation

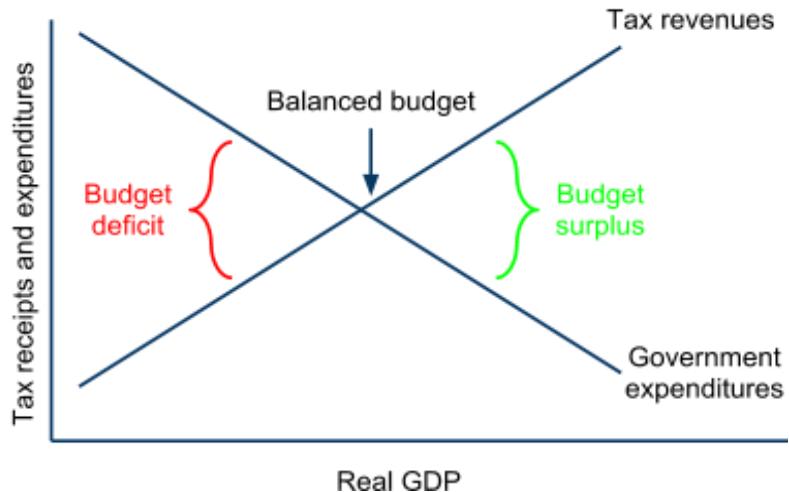
Next consider what happens when an economy already producing at full employment experiences an increase in investment spending.

- As AD increases, workers who were structurally or frictionally unemployed will be hired by firms producing capital goods, causing unemployment to fall below its natural rate.
- Newly employed workers start paying income tax, automatically increasing tax revenues for the government.
- Since they are no longer unemployed, the same workers will stop collecting unemployment benefits from the government automatically reducing the amount government pays in transfers.

During an expansion tax revenues increase automatically as GDP rises, slowing consumption and preventing the economy from overheating. At the same time, transfer payments decrease, helping prevent consumption from rising more than it otherwise would. The effect of these automatic stabilizers is to prevent an economy from overheating that otherwise might follow an increase in aggregate demand.

GDP, automatic stabilizers, and the government budget

The graph below shows the relationship between the level of GDP, tax revenues, and government expenditures.



Notice from the graph that when GDP decreases:

- Tax revenues automatically decrease because incomes fall when GDP falls. At lower incomes levels households pay less in tax.
- Government spending increases automatically as more households receive government welfare payments and more workers receive government unemployment benefits

Also notice that when GDP increases:

- Tax revenues increase because households' incomes and firms revenues increase. Some households move up to higher tax brackets and pay higher rates.
- Government spending decreases because fewer households depend on government support.

Evaluation of fiscal policy

- Evaluate the effectiveness of fiscal policy through consideration of factors including the ability to target sectors of the economy, the direct impact on aggregate demand, the effectiveness of promoting economic activity in a recession, time lags, political constraints, crowding out, and the inability to deal with supply-side causes of instability.

Time lags in fiscal policy

While fiscal policy can be a powerful tool for stimulating or contracting the level of aggregate demand and thereby influencing output, employment, and the price level, in reality it has some pretty major shortcomings. One in particular is that the implementation of fiscal policies is often hindered by **time lags**.

Often, by the time policymakers have identified a macroeconomic problem (a recession or an overheating economy), quantified it (determined the size of the output gap), determined the size of the multipliers and debated in congress or parliament the appropriate response,

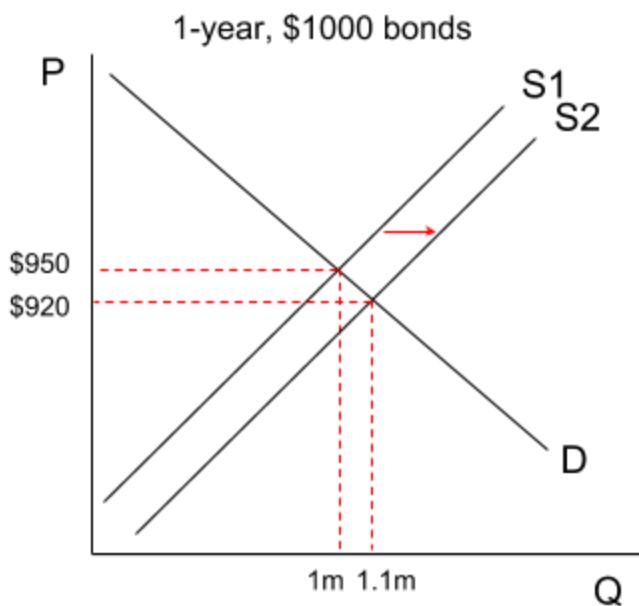
an economy will have already self-corrected or demand may have recovered on its own. A real economy is a dynamic machine that is always changing. Time lags refer to the fact that in the time it takes policymakers to figure out how to respond, everything could have changed in the economy.

For this reason, some government policymakers prefer to wait for an economy to self-correct from recessions or from excessive demand. The laissez faire, or “neo-classical” approach to economic policy is often to “wait and see” how an economy will respond to macroeconomic shocks such as demand and supply shocks.

The cost of deficits

Borrowing funds from the public to finance budget deficits comes with an opportunity cost for government. As with everything in life, there's no such thing as free debt. The bond market is where the government supplies bonds (certificates of debt) to investors, who buy and sell bonds and determine the price and the yield (interest rate) a government must pay its lenders.

As a government deficit spends, it must supply more bonds to the investors that lend it money. The graph below shows the effect on bond prices and the interest rate of increasing the supply of government bonds to finance a growing deficit.



The interest rate on a bond is inversely related to the bond's price. Before the government supplies more bonds (S₂), the interest rate the government must pay its lenders can be calculated:

- Investor's purchase price = \$950
- Price government pays after one year = \$1000
- Investor's return = \$50
- Interest on investment = $\frac{50}{950} \times 100 = 5.26\%$

Once the government increases the supply of bonds, which it must do to finance a growing deficit, the interest rate the government must pay investors increases:

- Investor's purchase price = \$920
- Price government pays after one year = \$1000
- Investor's return = \$80
- Interest on investment = $\frac{80}{920} \times 100 = 8.69\%$

As a government's total debt increases, it must pay interest on that debt just to maintain confidence among lenders that the government remains a safe entity to loan money to.

A government's total interest payments in a year can be roughly estimated by multiplying the total amount of debt by the interest rate expressed as a percentage. For example, assume the US government has \$20 trillion in debt. If interest rates on government bonds are currently 2.5%, then the United States must pay 2.5% of its \$20 trillion in debt to lenders in interest payments each year.

$$\text{Interest payments} = \text{Total debt} \times \text{interest rate (in hundredths)}$$

$$= \$20 \text{ trillion} \times 0.025 = \$500 \text{ billion}$$

The chart below shows the United States government's interest payments on public debt from 1996 to 2016.



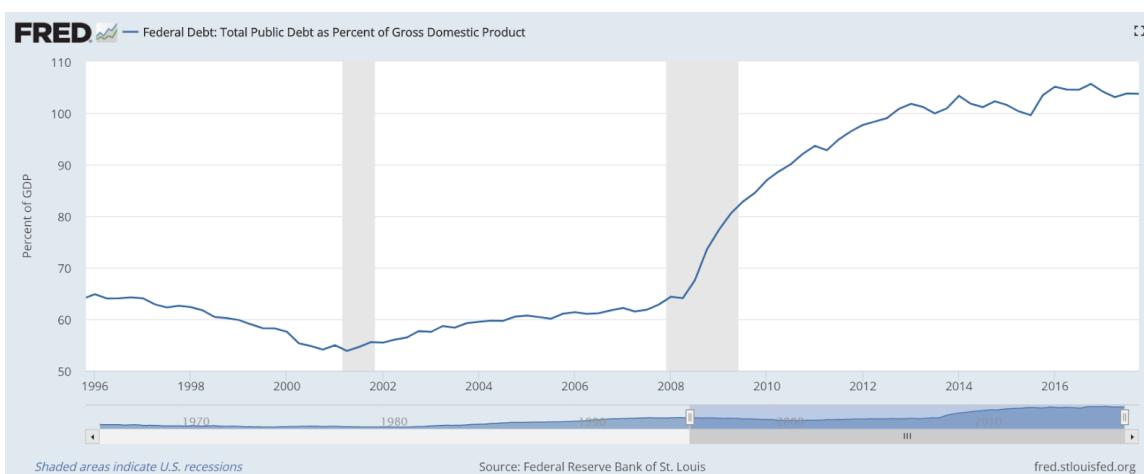
Notice that as the total debt increased over 20 years, interest payments increased to around \$500 billion in 2016.

The increased deficits and debt resulting from expansionary fiscal policy come with an opportunity cost: a greater amount of government spending must be allocated towards interest payments each year, potentially crowding-out other essential programs, goods, and services that government must pay for.

Debt to GDP ratio

While America's \$20 trillion debt sounds like a lot of money (and it IS!!) what really matters when considering the size of a country's debt is the ratio of its total debt to its total GDP. If any other country owed its lenders \$20 trillion, the burden of the interest payments alone would likely wipe out the government's budget and bring the economy to a screeching halt. However, while America's debt is big, its GDP has been even bigger for most of the last 20 years.

The chart below shows the percentage of the US GDP the total public debt made up from 1996 to 2016.



From 1996 to 2008, US public debt was about 60% of the country's national income. While this is not small, it is not unmanageable. Consider the example of our friend with her \$50 of weekly allowance from her parents. A debt burden of 60% for her would mean she owes her parents \$30 from all her past deficits. \$30 is not so much compared to her total income (allowance) of \$50. She can likely sustain \$30 of debt indefinitely in the future.

However, by 2012 the US public debt had grown to 100% of GDP. In other words, America's total debt to domestic and international lenders is now equal to the country's total income in a year (which means the US produced roughly \$20 trillion of total output in 2016).

Once a country's total debt exceeds 100% of its GDP, fiscal policymakers need to really consider the future costs of further increases in spending or decreases in taxation. As total debt grows, total debt payments grow as well, and the interest rate itself must increase to attract lenders willing to put their faith in the government's ability to repay its debts as the size of the country's debt grows relative to the country's income.

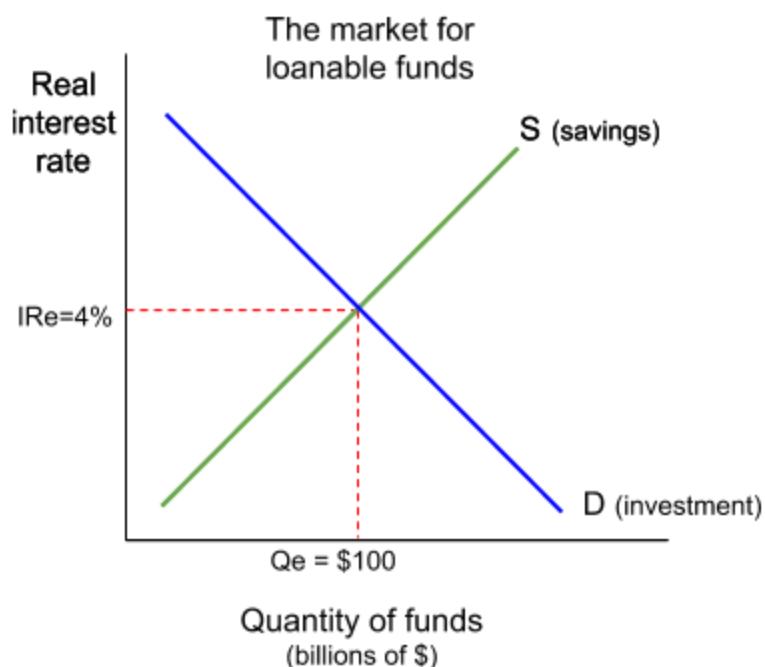
The increase in interest rates that often accompanies a government's deficit spending can create problems for private sector borrowers in a country (e.g. households and firms). This concept is explored further in the next section of this chapter.

The crowding-out effect

As explained in the previous section, when a government runs a budget deficit, it must borrow to finance its spending. The effects of government borrowing on the rest of the economy (households and firms) can be illustrated in the loanable funds market diagram.

The market for loanable funds shows the supply of household savings and the demand for investment from households and firms. The quantity of savings supplied increases as the real interest rate increases, while the quantity of funds demanded for investment decreases as interest rates increase.

Assume that in the market below, the demand for loanable funds represents only the demand from private borrowers (meaning households and firms). In other words, the government does not demand any funds from the private sector. The assumption then is that the government has a balanced budget, and finances all its spending through tax receipts.

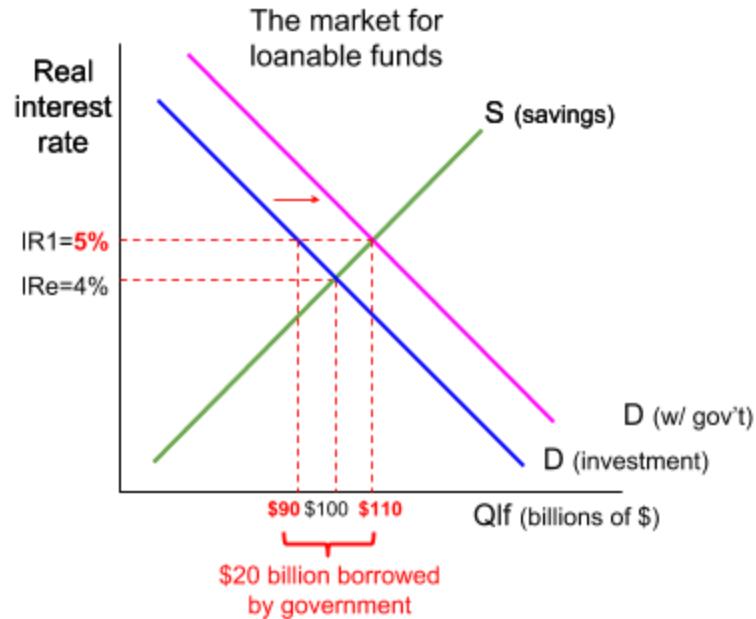


With no government borrowing, the equilibrium interest rate is 4% and the quantity of private sector investment and savings is \$100 billion.

What happens if the government runs a budget deficit and must borrow to finance its spending? The government becomes a demander of loanable funds, increasing the demand for funds and driving up the equilibrium interest rate.

Assume the government's budget deficit equals \$20 billion, which it must now borrow from the public. The graph below shows the effect of the government's \$20 billion deficit on the

equilibrium interest rate and the private quantity of funds demanded.



The government has entered the market for loanable funds, increasing the demand and driving the interest rate up to 5%. The effect the government's borrowing has on the private sector is can be observed:

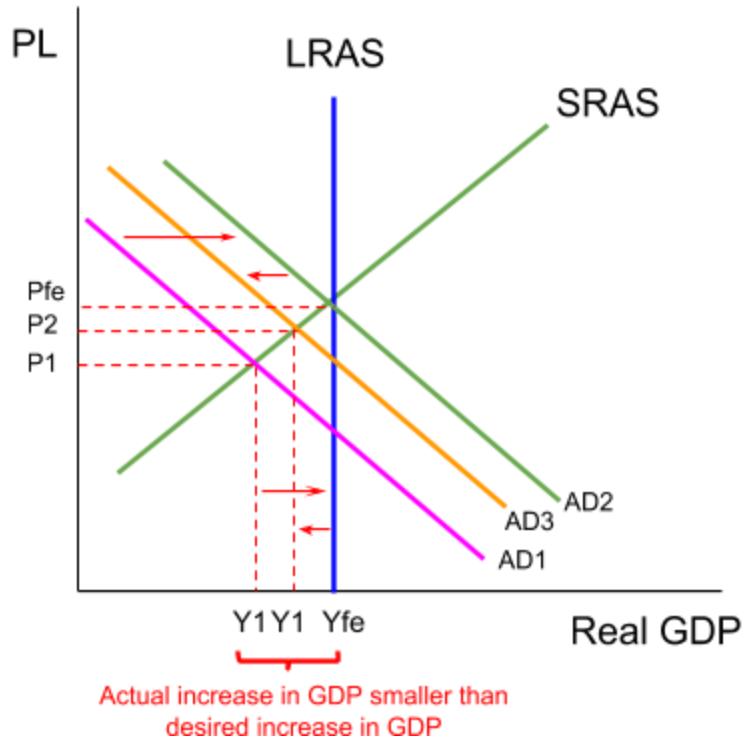
- Higher interest rates lead households to save more, increasing the quantity of funds supplied to \$110 billion.
- Higher interest rates lead to a fall in private sector investment (capital purchases by firms and investment in housing and real estate by households), decreasing the quantity of funds demanded to \$90 billion.
- Overall investment has increased by \$10 billion to \$110 as a result of the government's \$20 billion budget deficit.
- Total aggregate expenditures increases by less than the increase in the government spending because private sector spending is crowded out by the effect of the government's deficit spending.

Crowding out refers to the adverse effect of increased government borrowing, which leads to decreased levels of interest-sensitive private sector spending in the short run. As we showed in the graph, while overall spending increases by \$10 billion, the government's \$20 billion deficit has "crowded out" private sector spending by \$10 billion.

Implications of the crowding out effect for policymakers

Government fiscal policymakers should take into account the likelihood that crowding out will occur when deciding whether or not to enact an expansionary fiscal policy. If a deficit financed tax cut or increase in spending causes an increase in interest rates and a decrease in private sector spending, then the **multiplier effect** of any stimulus will be smaller than predicted by the simple spending multiplier formula.

In the graph below we can see the effect of an expansionary fiscal policy in a country experiencing a recessionary gap, assuming the increase in the budget deficit resulting from the fiscal action causes crowding out.



The government's fiscal stimulus intended to increase AD from AD1 to AD2. However, because higher interest rates **crowded out** private sector spending, the actual increase in AD was from AD1 to AD3. As a result, the actual increase in GDP is smaller than the intended increase in GDP and the economy is still in recession.

Impact of crowding out on long-run economic growth

Perhaps the most harmful effect of crowding out is the effect it could have on a country's long-run level of economic growth.

Recall that the primary source of economic growth is the increase in the quantity or the quality of the factors of production (land, labor, capital). Private sector investment is a major source of a country's capital stock. The **capital stock** is a country's accumulated supply of capital equipment and technology that is used in the production of goods and services. In other words, it's the tools and machinery needed to produce stuff that counts towards GDP.

If a government's budget deficit drives up interest rates and crowds out private investment in capital equipment and technology, then over time the capital stock in a nation will increase more slowly than it would otherwise. Crowding out can slow the long-run economic growth rate of a nation if investment in productivity-enhancing technologies is replaced with tax

cuts or government spending in less productive technologies or infrastructure.

Evaluating whether crowding out is likely to occur

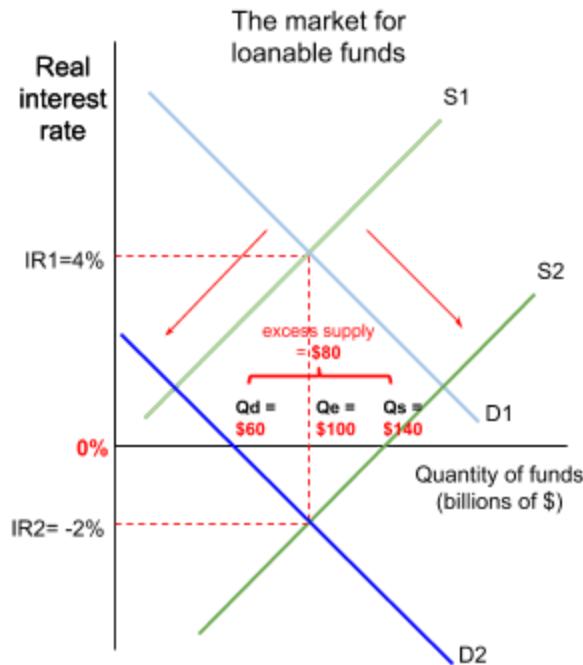
A country's government is most likely to engage in expansionary fiscal policies only when the economy is producing below its full employment level of output. During recessions, crowding out is less likely to be a concern than when fiscal stimulus is applied in an economy already producing at full employment.

During recessions, the private demand for loanable funds is already relatively low, and the supply is already high.

- Due to low consumer and business confidence, investment demand is weak during recessions.
- Due to uncertainty about future employment opportunities and the expectation of disinflation or deflation (lower prices in the future), household savings is higher during recessions (meaning the supply of loanable funds is greater).

If demand is very weak and supply is very high, a country may experience what's called a **liquidity trap** in its loanable funds market. A liquidity trap occurs when the equilibrium interest rate actually falls below 0%, but because banks are unwilling to offer negative interest rates, 0% acts as a price floor in the loanable funds market, creating an excess supply of loanable funds.

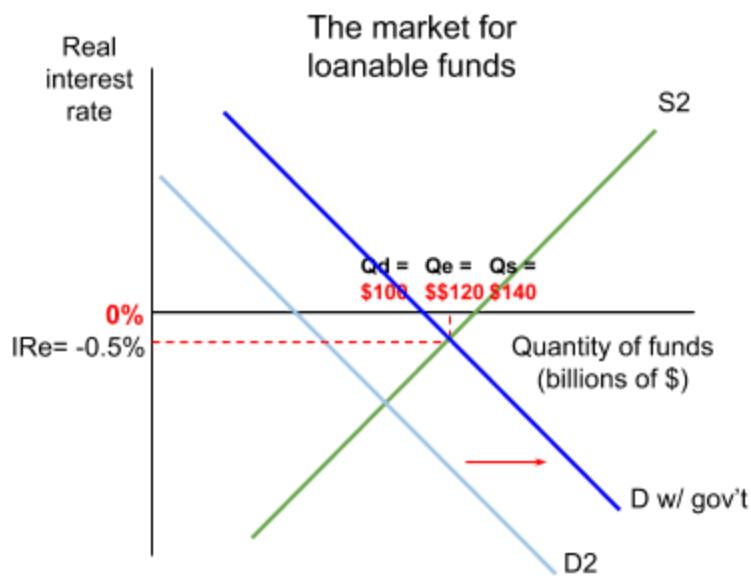
The graph below shows what may happen during a deep recession, when a liquidity trap is created in the loanable funds market.



In the graph above we can see what could happen in the loanable funds market during a deep recession.

- D1 and S1 represent the market when the economy is at full employment. The equilibrium interest rate is 4% and \$100 billion are supplied for investment.
- As the economy falls into recession, savings increases to S2 and investment decreases to D2. The new equilibrium interest rate is -2%, at which \$100 billion is supplied for investment.
- However, because interest rates are unlikely to fall below zero, 0% acts as a price floor in the market. At 0% only \$60 billion is demanded for investment, but households are willing to supply \$140 billion. There is an **excess supply of loanable funds**.

When the amount of savings exceeds the amount of investment by the private sector, as seen in the graph above, there is room for government to borrow from the public without resulting in crowding out. Let's show the effect of increased government borrowing in a loanable funds market experiencing a liquidity trap:



In the graph above we see the effect of a government's borrowing to finance a budget deficit during a deep recession:

- Demand for funds has increased to D w/gov't, causing the equilibrium interest rate to increase from -2% to -0.5%.
- 0% still acts as a price floor in the market, so the quantity of savings (\$140 billion) still exceeds the quantity of investment (\$100 billion).
- However, there has been no decrease in private sector investment because of the government's intervention. Overall investment has increased from \$60 billion to \$100 billion with no fall in private sector spending, because there has been no change in the actual interest rate, which remains at 0%.

- The economy enjoys a \$40 increase in overall spending without any fall in private spending.

During deep recessions, when interest rates are already extremely low (close to 0%), crowding out is much less likely to occur because there is already an excess supply of loanable funds. Government can borrow excess funds that otherwise would remain uninvested without driving up interest rates.

In summary, when the private sector is underinvesting and oversaving, government can deficit spent and stimulate aggregate demand through expansionary fiscal policies without worrying about driving up interest rates and crowding out private sector investment. Basically, when there's very little private sector investment happening, crowding out is of little concern to fiscal policymakers.

2.5 Monetary Policy

Interest rate determination and the role of a central bank

- Describe the role of central banks as regulators of commercial banks and bankers to governments.
- Explain that central banks are usually made responsible for interest rates and exchange rates in order to achieve macroeconomic objectives.
- Explain, using a demand and supply of money diagram, how equilibrium interest rates are determined, outlining the role of the central bank in influencing the supply of money.

While fiscal policy is enacted by a country's government with the aim of managing the level of aggregate demand to promote certain macroeconomic objectives, monetary policy is enacted by a country's **central bank**, which is a national bank that provides regulatory and banking services to a country's commercial banks and other lending institutions.

Monetary policy involves a central bank increasing or decreasing the nation's money supply to influence interest rates and the level of aggregate demand with the aim of promoting macroeconomic objectives such as price level stability and full employment.

The **nominal interest rate** is the annual rate a borrower must pay back a lender for the use of borrowed money. Nominal interest rates are determined in the money market, where the supply of money is determined by a country's central bank policy and the demand for money is determined by the nation's households, firms, government, and foreigners; basically, anyone who needs money as an asset or to purchase goods and services.

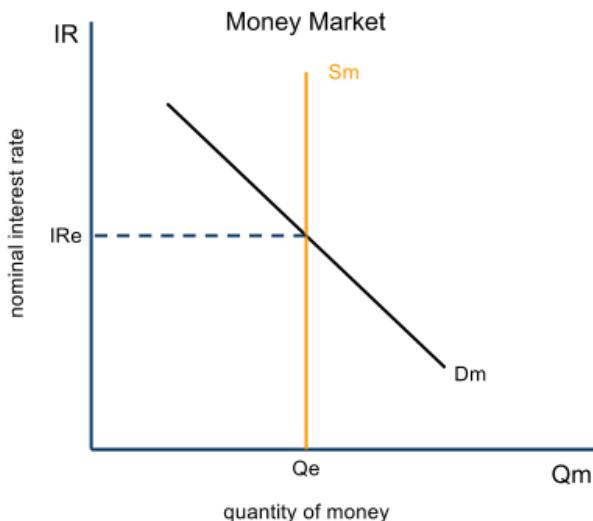
The **money market** shows the relationship between nominal interest rates and the supply and demand of money in an economy.

Money demand represents the quantity of money (M_1) the public demands at a range of interest rates as an asset and for the purchase of goods, services, and resources. There are two elements of a nation's money demand:

- **Asset demand:** The demand for money as an asset is inversely related to the interest rate. As explained earlier in the chapter, at high interest rates less money is demanded, because the opportunity cost of holding money as an asset is higher. At low interest rates the opportunity cost of holding money as an asset decreases and the quantity demanded is higher.
- **Transaction demand:** The transaction demand for money depends on the level of output produced in the nation and the interest rate. At lower interest rates households are more willing to spend money on goods and services, while at higher interest rates the public demands less money for transactions, since the opportunity cost of buying stuff is higher when more interest can be earned in financial assets.

Money supply is independent of the nominal interest rate, and determined by the country's central bank. A central bank can increase or decrease the supply of money and thereby change the nominal interest rate through the use of **monetary policy**. Because it is determined by the central bank, money supply is not responsive to changes in interest rates. In other words, it's perfectly inelastic.

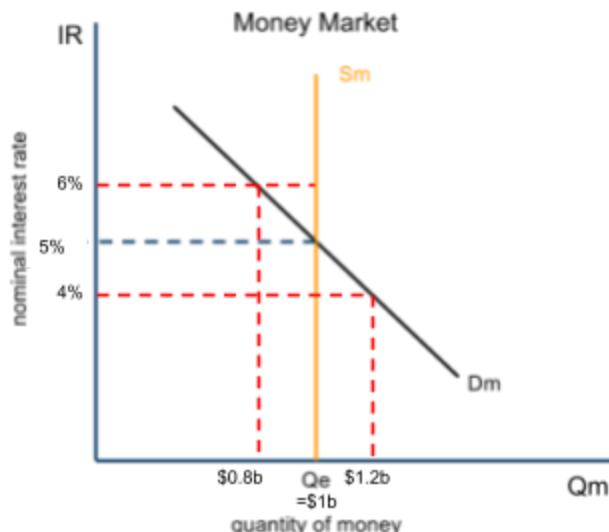
The graph below shows the money market in a nation:



Equilibrium and disequilibrium in the money market

The **equilibrium interest rate** is determined by the supply of and demand for money in a country. A shift in the demand for money can lead to a change in the equilibrium interest rate. At equilibrium the quantity of money supplied is equal to the quantity demanded.

A money market will be in disequilibrium when the quantities demanded and supplied are not equal.



In the money market above, we can see why 5% is the equilibrium interest rate.

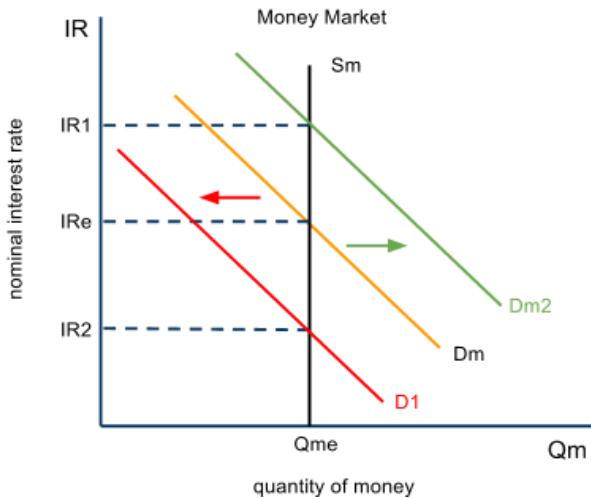
- **At 6%** the quantity of money demanded as an asset or for transactions by households, firms, the government, and foreigners is only \$0.8 billion. The quantity supplied by the central bank, however, is \$1 billion. There is a surplus of money in the economy, most likely sitting in banks' excess reserves going unloaned. Interest rates must fall in order for the money market to clear.
- **At 4%** the quantity of money demanded as an asset or for transactions is \$1.2 billion, more than the \$1 billion actually supplied. There is a shortage of money in the economy, meaning there is greater demand for money for spending and investments than there is supply. Interest rates must rise for the money market to clear.
- **At 5%** the quantity demanded equals the quantity supplied. Banks are mostly loaned out and do not have lots of excess reserves. Nor are there many borrowers who are unable to get money for the spending they would like to do.

Changes in money demand

Money demand will shift whenever there is a change any of the following:

- **Real GDP:** An increase in real GDP will increase income and consequently the demand for money throughout the economy. A fall in GDP causes money demand to decrease as there are fewer goods and services to buy.
- **Price level:** A higher price level will lead to higher demand for money as more money will be required to buy a given set of goods and services. A fall in prices will cause demand for money to decrease as less money is needed to buy the same amount of stuff.
- **Expectations about future price levels:** Even the expectation of higher prices in the future can increase demand for money today, as households and firms will rush to buy the things they need now before prices rise. Expected deflation will decrease demand for money as consumers and firms postpone spending today and wait for prices to fall in the future.

The graph below shows the effect of a change in money demand on the equilibrium interest rate in the money market.



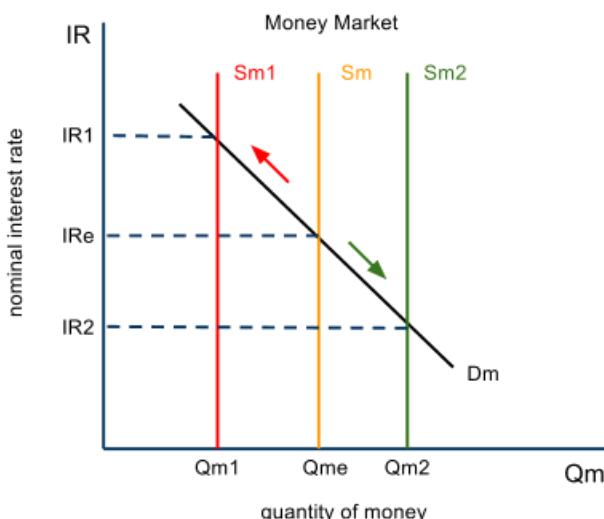
Observe the following:

- An increase in GDP or an increase in inflation causes an increase in demand for money to D_m , which makes money scarcer. Banks must raise interest rates as money demand rises in order to prevent their being shortages of funds.
- A decrease in GDP or a decrease in the price level causes a decrease in demand for money to D_1 , which makes money less scarce. Banks must lower interest rates as money demand decreases in order to avoid an unwanted increase in their excess reserves.

Changes in the money supply

If the supply of money changes, the equilibrium interest rate will change in the economy. Money supply changes result from monetary policy actions taken by the central bank.

The graph below shows the effect of shifts in the money supply curve on the equilibrium interest rate.



Central bank action that causes the money supply to increase (from Sm to $Sm2$) will cause interest rates to fall. Banks' excess reserves increase, so they will lower the rates they charge

to attract more borrowers.

Central bank action that causes the money supply to decrease (from S_m to S_{m1}) will cause interest rates to rise. Banks' have less in their excess reserves and some may even have to raise rates to attract more depositors or and fewer borrowers. Unless rates increase, banks will fall short in meeting their required reserves and there will be a shortage of money.

Monetary policy and short-term demand management

- Explain how changes in interest rates can influence the level of aggregate demand in an economy.
- Describe the mechanism through which easy (expansionary) monetary policy can help an economy close a deflationary (recessionary) gap.
- Construct a diagram to show the potential effects of easy (expansionary) monetary policy, outlining the importance of the shape of the aggregate supply curve.
- Describe the mechanism through which tight (contractionary) monetary policy can help an economy close an inflationary gap.
- Construct a diagram to show the potential effects of tight (contractionary) monetary policy, outlining the importance of the shape of the aggregate supply curve.

Introduction to monetary policy

A central bank's manipulation of the money supply and nominal interest rates is known as **monetary policy**. Central banks implement monetary policies to achieve macroeconomic goals, such as price stability, full employment, and economic growth.

A **central bank** is the institution in most modern, market economies that controls the overall supply of money in the nation's economy. Most central banks act independently of the nation's government, and are thus, in theory, insulated from political agendas and influence. Examples include:

- In the US: the Federal Reserve Bank
- In the UK: the Bank of England
- In China: the People's Bank of China
- In Japan: the Bank of Japan
- In Switzerland: the Swiss National Bank
- In the Eurozone: the European Central Bank

Every major world economy has a central bank. Below is a snapshot of one CB and the roles it plays in the nation's banking system and wider economy

The Federal Reserve Bank of the United States	
Overview of the Federal Reserve Bank of the United States	<ul style="list-style-type: none"> • 12 branches located in different regions of the country • Coordinated by the Fed's Board of Governors • The "Fed" provides banking services to commercial banks <ul style="list-style-type: none"> ➢ Accept deposits, lends money (called the "discount window", only if commercial banks can't borrow from one another would they borrow from the Fed), issues new currency to private banks • FOMC - Federal Open Market Committee: 12 individuals, including the Chairman of the Fed (Bernanke). Purpose is to buy and sell government securities to control the nation's money supply and influence interest rates. Execute monetary policy.
Functions of the Federal Reserve Bank	<ul style="list-style-type: none"> • Issue currency: the Fed can inject new currency into the money supply by issuing Federal Reserve Notes (dollars) to commercial banks to be loaned out to the public. • Setting reserve requirements: this is the fraction of checking account balances that commercial banks must keep in their vaults. The larger the reserve requirement, the less money commercial banks can loan out. • Lending money to banks: The Fed charges commercial banks interest on loans, this is called the "discount rate". • Controlling the money supply: this in turn enables the Fed to influence interest rates.

The tools of monetary policy

Changing the money supply will cause interest rates to increase or decrease, which can then influence the level of aggregate expenditures in the economy. A central bank has three tools for increasing or decreasing the supply of money in an economy:

- **The buying and selling of government bonds:** Every commercial bank will invest some of its depositors money in illiquid government bonds. Bonds are not money, so if a central bank wishes to increase the supply of money in the economy, it can buy bonds from commercial banks using newly printed cash (which IS money!) If the goal is to reduce the money supply, a central bank can sell bonds to commercial banks, which results in less money in circulation and more illiquid government bonds on banks' balance sheets.
- **Changing the required reserve ratio:** The required reserve ratio (RRR) is the percentage of a bank's total deposits it is required to keep in reserve. By reducing the reserve requirement, a central bank immediately increases commercial banks' excess reserves, which frees up money for new loans. By increasing the reserve requirement, a central bank immediately reduces the amount of excess reserves in the banking system and commercial banks must raise interest rates to meet the higher reserve requirement.
- **Changing the discount rate:** The discount rate is the interest rate the central bank charges commercial banks for short-term loans. If this rate is lowered, banks will be more willing to make loans to private borrowers and interest rates will fall. If the discount rate is increased, banks will be less willing to loan to private borrowers and the interest rate will increase.

Next we'll go into more detail of how each of the tools of monetary policy works.

Buying and selling of bonds

Sometimes called “**open market operations**”, a central bank’s interventions in the bond market is the most commonly employed monetary policy tool. Open market operations can be employed as either an **expansionary monetary policy** (one that increases the money supply and reduces interest rates) or a **contractionary monetary policy** (one that reduces the money supply and increases interest rates).

When the goal is to reduce interest rates and stimulate aggregate demand, a central bank will buy bonds from commercial banks and the public. An open market purchase of government bonds will cause the money supply to increase by a magnitude determined by the **money multiplier**.

For example, assume the central bank of Wahoovia desires to reduce interest rates and to do so it aims to increase the money supply by \$10 billion. The reserve requirement is 20% in Wahoovia. The money multiplier can be calculated:

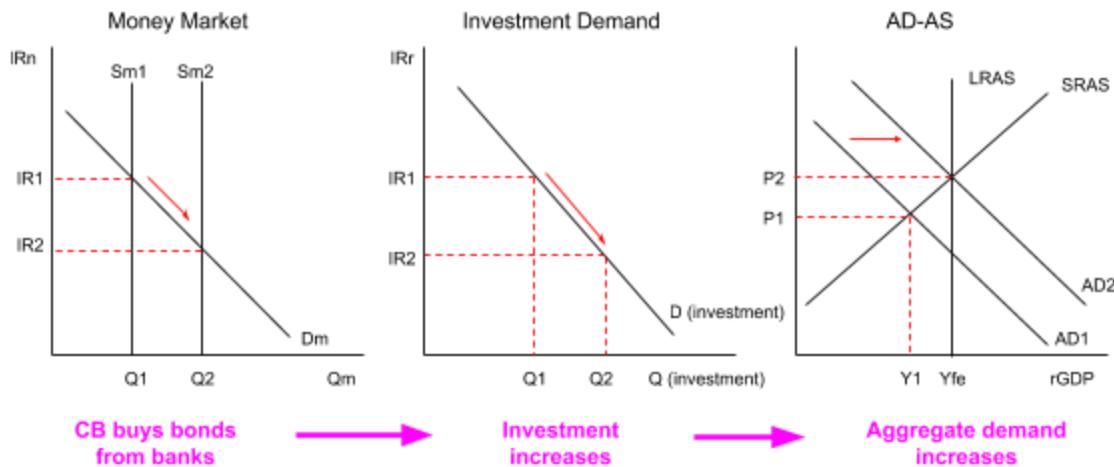
$$\begin{aligned}\text{Money multiplier} &= \frac{1}{\text{RRR}} \\ &= \frac{1}{0.2} = 5\end{aligned}$$

If the central bank wishes to increase the money supply by \$10 billion, it must purchase \$2 million in government bonds from the public. Doing so will increase banks’ excess reserves by \$2 billion, which will increase the money supply based on the money multiplier.

$$\begin{aligned}\Delta \text{ in money supply} &= \Delta \text{ in excess reserves} \times \text{money multiplier} \\ &= \$2 \text{ billion} \times 5 = \$10 \text{ billion}\end{aligned}$$

A \$2 billion purchase of government bonds by the central bank of Wahoovia will increase the money supply by \$10 billion. Banks will loan out the initial \$2 billion increase in their excess reserves, which will create new deposits and new loans across the banking system until the ultimate increase in the money supply is multiplied five times.

The effect of an expansionary monetary policy can be observed in the graphs below:



The purchase of government bonds by the central bank has resulted in a fall in the nominal and the real interest rate, an increase in investment (and interest sensitive consumption such as the purchase of new cars) and an increase in AD.

- Notice that before the expansionary monetary policy this country had a recessionary gap of $Y_1 - Y_{fe}$.
- However, after the stimulus, AD has increased to the full employment level.

A contractionary monetary policy will have the opposite effect on output, employment, and the price level. Assume that rather than a recession, Wahoovia is facing high inflation.

An open market sale of government bonds by the central bank will reduce the money supply, raise interest rates, and reduce interest sensitive spending in the economy.

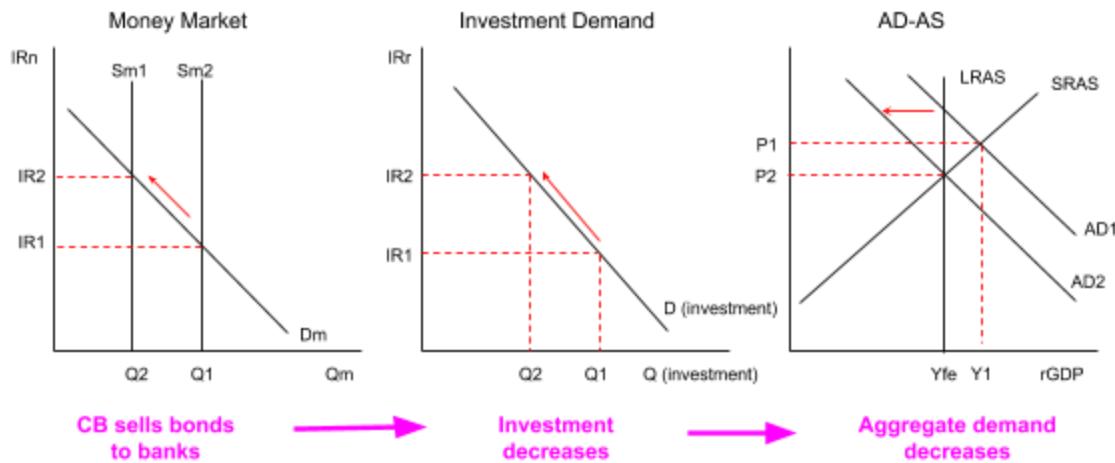
Let's assume the central bank wishes to reduce the money supply by \$5 billion. With the money multiplier, we can calculate how much the central bank should sell in government bonds to the public. All we must do is divide the desired change in the money supply by the money multiplier:

$$\text{needed sale of bonds} = \frac{\text{desired change in money supply}}{\text{money multiplier}}$$

$$= \frac{\$5 \text{ billion}}{5} = \$1 \text{ billion}$$

The central bank of Wahoovia should sell \$1 billion in government bonds to the public. Doing so will reduce the supply of liquid money in the economy, reducing banks' excess reserves and reduce the overall money supply by \$5 billion.

The effect of a contractionary monetary policy can be observed in the graphs below.



The sale of government bonds by the central bank has resulted in an increase in the nominal and the real interest rates, a decrease in investment (and interest sensitive consumption such as the purchase of new cars) and a decrease in AD.

- Notice that before the expansionary monetary policy this country had an inflationary gap of $Y_{fe} - Y_1$.
- However, after the stimulus, AD has fallen to the full employment level.

The reserve ratio as a tool of monetary policy

Changing the reserve ratio is a powerful way to stimulate or reduce total spending in the economy. It impacts more than just the proportion of deposits banks must keep in reserve.

For example, assume the US Fed wishes to reduce the total amount of money in circulation to increase the interest rate and reduce consumption and investment. By raising the reserve ratio, it can achieve a smaller money supply and a higher interest rate, but also a smaller money multiplier.

The table below shows the effect of an increase in the RRR from 0.10 to 0.15.

	Before the Fed's action	After the Fed's Action
Required Reserve Ratio	0.10	0.15
Money Multiplier	$\frac{1}{0.10} = 10$	$\frac{1}{0.15} = 6.67$

Effect of the Fed's Action:

- With fewer excess reserves to lend out, the money supply decreases and the interest rate rises.
- When new deposits are made, banks must now keep a larger proportion in reserve, reducing the overall money supply in the economy.

- For every \$1 increase in excess reserves in the future, only \$6.67 of new money will be created compared to \$10 before the Fed's action.

Changing the reserve requirement is a powerful, albeit rarely used tool of contractionary monetary policy.

Reducing the reserve requirement increases the money supply and could be used as an expansionary monetary policy.

- A decrease in the reserve requirement from 0.10 to 0.5 would double the money multiplier from 10 to 20.
- Banks would immediately see their required reserves halve, increasing their excess reserves and the amount of new loans they could make in the economy.
- For every \$1 increase in excess reserves in the future, \$20 of new money could be created compared to just \$10 before the Fed's action.

The discount rate as a tool of monetary policy

The discount rate is the interest rate that the central bank charges to commercial banks that wish to borrow funds to meet shortfalls in their required reserves. The central bank will make short term loans to commercial banks if they wish to make loans that would otherwise result in their reserves falling below the required level.

The central bank is the **lender of last resort** for commercial banks, meaning that generally commercial banks prefer to borrow from one another to meet their reserve requirements, but when not enough funds are available from other commercial banks, they can turn to the central bank to meet shortfalls in their required reserves.

By increasing the discount rate, the central bank sends a signal to commercial banks that making loans beyond what they have in their excess reserves is a bad idea, because the cost of repaying the borrowed funds will be higher for the bank.

Lowering the discount rate sends the signal that it is okay to make loans beyond what commercial banks have in their excess reserves, because borrowing from the central bank to make up the shortfall is relatively cheap!

Like the reserve requirement, changing the discount rate is a relatively infrequently used tool of monetary policy.

The relative importance of the three monetary policy tools

The three tools of monetary policy are called into action to varying degrees by the world's central banks. The most commonly used tool is open market operations, while reserve ratios and discount rates tend to be changed less frequently.

Relative Importance of the Monetary Policy Tools	
Open Market Operations	Open-market operations are the buying and selling of government bonds in the financial market. Because it is the most flexible, bond holdings by the central bank can be adjusted daily, and have an immediate impact on banks' reserves and the supply of money in the economy
Reserve Ratio	The required reserve ratio is RARELY changed. RRR in the US has been .10 since 1992. Reserves held by the Central Bank earn little or no interest; therefore if RRR is raised, banks' profits suffer dramatically since they have to deposit more of their total reserves with the Fed where they earn almost no interest. Banks prefer to be able to lend out as much of their total reserves as possible
Discount Rate:	Until recently, the discount rate in the US was rarely adjusted on its own, and instead hovered slightly above the federal funds rate. In 2008, the US Fed lowered the discount rate to very low levels as uncertainty among commercial banks brought private lending to a halt. The "discount window" is only supposed to be used in the case of private lenders being unable to acquire funds, hence the Fed is the lender of last resort

Monetary policy and inflation targeting

- Explain that central banks of certain countries, rather than focusing on the maintenance of both full employment and a low rate of inflation, are guided in their monetary policy by the objective to achieve an explicit or implicit inflation rate target.

As we have demonstrated above, changes in the money supply can have immediate impacts on aggregate expenditures, output, employment, and the price level. In this way, monetary policy provides a powerful tool for stimulating or contracting the overall economy and promoting macroeconomic stability.

The primary objective of some central banks, however, is to focus their policies on promoting price level stability; more specifically, central banks tend to have a **target inflation rate** that they attempt to maintain in the economy. For most major, developed economies, the target inflation rate is typically in the range of 2% to 3%.

If actual inflation is below the target rate, central banks tend to engage in expansionary monetary policy. For example, assume Wahoovia is currently experiencing inflation of just 1%.

- The central bank fears that this low rate of inflation could discourage investment and possibly lead to deflation and rising unemployment.
- The central bank should buy bonds on the open market.
- The money supply increases and interest rates fall.
- Low interest rates will make borrowing and spending more attractive to businesses

and households.

- More investment and consumption should boost AD and increase the equilibrium price level.
- When inflation returns to the 2%-3% range the central bank could stop buying bonds. Inflation is back within its target range.

What if inflation is higher than the target rate? Say Wahoovia instead has 5% inflation.

- The central bank fears this will lead to even higher inflation as consumers scramble to buy things before they get even more expensive.
- Also, high inflation is eroding the real income of Wahoovians.
- The central bank decides to intervene and sell bonds on the open market.
- The money supply decreases and interest rates rise.
- Higher interest rates make borrowing more costly so households and firms prefer to save more and spend less.
- Less investment and consumption reduces AD and brings down the equilibrium price level.
- When inflation returns to the 2%-3% range the central bank could stop selling bonds. Inflation is back within its target range.

Evaluation of monetary policy

- Evaluate the effectiveness of monetary policy through consideration of factors including the independence of the central bank, the ability to adjust interest rates incrementally, the ability to implement changes in interest rates relatively quickly, time lags, limited effectiveness in increasing aggregate demand if the economy is in deep recession and conflict among government economic objectives.

As with fiscal policy, the effectiveness of monetary policy may be hindered by **time lags**. By the time central bank policymakers have identified and quantified a macroeconomic problem (deflation, high inflation), and intervened to correct the problem, there is the chance that macroeconomic conditions could have changed sufficiently to render the bank's intervention ineffective.

Additionally, the effectiveness of monetary policy as a tool for combating inflation or a recession may be limited by other factors.

Factors that may limit the effectiveness of monetary policy	
The degree of inflation:	<p>In periods of extremely high inflation, it is unlikely that a contractionary monetary policy alone will be adequate to bring inflation under control.</p> <ul style="list-style-type: none"> ➤ The expectation of high inflation creates a strong incentive among households and firms to spend money in the present rather than waiting till the future, when prices are expected to be higher. ➤ A substantial increase in interest rates (to a level higher than the expected inflation rate) would be required to reign in present spending reduce aggregate demand ➤ Contractionary fiscal policy (higher taxes, reduced government spending) may be needed to support higher interest rates during periods of high inflation

The depth of the recession:	<p>In periods of weak demand, high unemployment and deflation, it is unlikely that an expansionary monetary policy alone will be adequate to bring an economy back to full employment</p> <ul style="list-style-type: none"> ➤ When private spending (consumption and investment) are deeply depressed, a decrease in interest rates may not be enough to stimulate spending and AD ➤ With the expectation of future deflation, the private sector has a strong incentive to save, since money saved now will be worth more in the future. ➤ Expansionary fiscal policy may be needed to reinforce the decrease in interest rates to boost demand to its full employment level.
------------------------------------	--

Supply-side effects of monetary policy

While monetary policy is generally considered a demand-side policy (since changes in interest rates directly affect investment, a component of AD), it can also impact the level of aggregate supply in a country.

For example, assume a country's central bank lowers interest rates to stimulate AD during a recession.

- Lower interest rates lead to more investment and consumption, so AD increases.
- More investment leads to an increase in the nation's capital stock.
- More capital makes labor more productive and reduces production costs over time, increasing SRAS and LRAS.
- The increase in AD brings the recession to an end, while the increase in LRAS means the economy's potential output has increased.

If the economic conditions are right and firms are willing to invest, expansionary monetary policy can contribute to long-run economic growth!

2.6 Supply-side Policies

Supply-side policies and the economy

- Explain that supply-side policies aim at positively affecting the production side of an economy by improving the institutional framework and the capacity to produce (that is, by changing the quantity and/or quality of factors of production).
- State that supply-side policies may be market-based or interventionist, and that in either case they aim to shift the LRAS curve to the right, achieving growth in potential output.

The role of government in promoting economic growth

In an earlier section of this chapter we learned how government can negatively impact the long-run economic growth of a country if its deficit spending crowds out private sector investment and slows the growth of the country's capital stock. However, there is plenty a government can do to promote long-run economic growth as well.

Throughout the course we have studied fiscal and monetary policies' effects on aggregate demand. These "demand-side policies" are effective at stimulating or contracting the amount of aggregate spending and promoting the short-run macroeconomic objectives of price level stability and full employment. We have not considered, however, the impact these and other policies can have on a nation's aggregate supply and its level of potential output in the long run.

Supply-side policies are measures undertaken by the government aimed at increasing the level of aggregate supply in a country, and thereby promote long-run economic growth.

Examples of government supply-side policies include:

- Reducing business and income taxes
- Labor market reforms
- Deregulation of industries
- Trade liberalization
- Investment in human capital
- Investment in physical capital

Supply-side policies increase productivity and reduce production costs, shifting SRAS and LRAS outwards, increasing actual and potential output.

Supply-side policies can be either interventionist or market-based. **Interventionist supply-side policies** involved government increasing its involvement in the economy, either through investments in human capital, investments in physical capital, infrastructure spending, or industrial policies. **Market-based supply-side policies** involved government decreasing its involvement in the economy, either through deregulation, tax reduction, trade liberalization or labor market reforms.

Interventionist supply-side policies - Investment in human capital

- Explain how investment in education and training will raise the levels of human capital and have a short-term impact on aggregate demand, but more importantly will increase LRAS.

Not all investment needed to promote economic growth must be undertaken by the private sector. Supply-side fiscal policies affect aggregate demand, aggregate supply, and potential output in the short run and long run by improving the quantity and the quality of physical capital and by improving human capital.

There are certain goods that are under-provided by the free market. Government provision or subsidizing of such goods can improve efficiency in an economy and increase the country's potential output. **Education** is such a good. Because the benefits of having an educated population are social in nature, it is often up to the government to promote education through public schools.

- A better educated workforce is more productive and provides firms with higher-skilled workers to produce more and better output.
- A better educated workforce will earn higher incomes and thus pay more in taxes, allowing government to earn more revenues with which it can provide other merit and public goods.

Interventionist supply-side policies - investment in new technology

- Explain how policies that encourage research and development will have a short-term impact on aggregate demand, but more importantly will result in new technologies and will increase LRAS.

Research and development (R&D) occurs whenever a private firm or the government spends money to develop new technologies or to promote the overall level of scientific knowledge in society. Government can support R&D through its spending on defense, aeronautical and space programs, scientific research, and other programs. The discoveries and technologies resulting from government-funded R&D have the potential to increase productivity or to create entire new industries, increasing the country's potential output and shifting LRAS outwards.

Interventionist supply-side policies - Investment in infrastructure

- Explain how increased and improved infrastructure will have a short-term impact on aggregate demand, but more importantly will increase LRAS.

Due to the fact that they are non-excludable, certain types of **infrastructure** (roads, sanitation, electricity grids, communication infrastructure) may need to be provided by the government

- Better infrastructure reduces the costs for private businesses and allows them to operate more efficiently.

- Because private firms do not have to build their own roads or railways, they are able to produce and sell their products at lower costs. A modern, efficient infrastructure allows for AS to increase over time.

Interventionist supply-side policies - Industrial policies

- Explain that targeting specific industries through policies including tax cuts, tax allowances and subsidized lending promotes growth in key areas of the economy and will have a short-term impact on aggregate demand but, more importantly, will increase LRAS.

Reducing business and personal income taxes can positively impact aggregate supply in two ways:

- Lowering production costs:** Lowering taxes on firms reduces the cost of doing business. Allowing business owners to keep a larger share of their earned revenues should incentivize new investments in capital and technology, which increase the productivity of labor and reduce costs, shifting AS outwards.
- Increasing the incentive to work:** One argument against high income taxes is that if you tax higher incomes at higher rates, households have a disincentive to work hard and earn higher incomes. In contrast, if tax rates are lowered households have a greater incentive to work, so labor force participation could increase and worker productivity could improve, increasing SRAS, LRAS, and potential output.

Market-based supply-side policies - increased competition

- Explain how factors including deregulation, privatization, trade liberalization and anti-monopoly regulation are used to encourage competition.

Deregulation: One of the many roles of government is to regulate the activities of producers in the free market. The goals of regulation are often to reduce the impact of negative impacts on the environment, human health, or society arising from the production or consumption of certain goods. The cost of complying to government regulation increases firm's average costs and reduces the level of aggregate supply. Removing or loosening regulations in certain industries will lead to lower costs and greater output, increasing aggregate supply.

For example, the United States Environmental Protection Agency is often accused of imposing harmful regulations on producers of goods ranging from automobiles to electricity to agricultural products.

- The benefit of environmental regulations is the reduced emissions of harmful toxins that affect human health and the environment.
- The cost of environmental regulations is the impact they have on employment and the price level.
- Reducing regulations will lead to more output and lower prices in the regulated industries, increasing actual and potential output in the country as a whole.
- However the tradeoff may be increased environmental degradation and reduced human health.

Trade liberalization is the removal of barriers to trade between a country and other countries. The reduction or removal of **protectionist policies** will allow a country's firms to acquire cheaper inputs from abroad and increase the potential market for a country's producers by opening the market to international consumers.

Protectionism refers to the use of tariffs, quotas or other measures aimed at making domestic producers more competitive with foreign producers by limiting the quantity of imports into the nation.

- **Tariffs** are taxes placed on imported goods, services or resources
- A **quota** is a physical limit on the quantity of a good, service or resource that may be imported

The supply-side effects of trade liberalization include:

- Reducing protectionism will allow producers in a country to import raw materials and other factors of production more cheaply, reducing average production costs
- More competition from foreign producers will force domestic firms to use their resources in a more efficient manner, since they will either have to reduce their production costs or lose out to foreign competition.

Both lower production costs and increased competition lead to an increase in the nation's aggregate supply and contribute to long-run economic growth.

Anti-monopoly legislation and other efforts by government to increase the degree of competition in the markets for consumer goods will increase the level of efficiency and competition in various industries. The entrance of more firms into product markets increases aggregate supply and promotes long-run economic growth.

Market-based supply-side policies - Labor market reforms

- Explain how factors including reducing the power of labour unions, reducing unemployment benefits and abolishing minimum wages are used to make the labour market more flexible (more responsive to supply and demand).

Labor is the most important (and the most costly) resource for most nations' producers. Reforms of the labor market that bring down the cost of labor will increase a nation's aggregate supply and lead to growth in national output. Supply-side labor market reforms include:

- **Reducing or eliminating the minimum wage:** The minimum wage is a price floor in the labor market set above the free market equilibrium wage rate. Minimum wage laws increase the cost of hiring workers in certain industries (typically the low-skilled sectors). Reducing or eliminating minimum wages may lead firms to hire more workers and thereby produce more output at a lower per unit cost
- **Reducing labor union power:** Labor unions are organizations of workers in particular industries that negotiate with employers for better worker benefits, such as higher wages, more paid vacation, better health care and so on. Such benefits add to

firms' production costs and keep aggregate supply lower than it might be otherwise. Reducing the power of unions will lower labor costs for producers and shift AS outwards.

- **Reducing government spending on unemployment benefits:** Unemployment benefits are the money payments individuals received during the period of time when they are out of work and seeking a new job. Reducing these benefits would create an incentive for unemployed workers to accept a new job more quickly and at a lower wage rate than they otherwise might accept. Firms will find more workers willing to work for lower wages.

Evaluation of supply-side policies

- Evaluate the effectiveness of supply-side policies through consideration of factors including time lags, the ability to create employment, the ability to reduce inflationary pressure, the impact on economic growth, the impact on the government budget, the effect on equity, and the effect on the environment.

The implementation of the supply-side policies described above can have a positive effect on a country's rate of long-run economic growth.

Fiscal Policy's positive supply-side effects	
Infrastructure spending:	When government supports a modern infrastructure, including for transportation and communications, the private sector is given the resources it needs to grow and succeed in the long-run
Education spending:	Human capital is perhaps the most important resource a nation requires for long-run economic growth. Public, government funded schools and programs to improve skills in the labor can contribute to long-run growth.
Research and development:	Government-funded research and development can lead to scientific, technological, and medical breakthroughs that may spur new industries and promote growth across the private sector.
Incentives for private investment:	Creating a tax policy that rewards innovation and entrepreneurship, rather than punishes it by taking the 'winners' in an economy will encourage private businesses to invest and thereby help the economy grow.

On the surface, public policies that increase aggregate supply and promote economic growth are overwhelmingly beneficial for an economy:

- The average price level of goods and services decreases, while...
- Output increase, and...
- Employment increases...

However, such policies are often difficult to undertake and may have some undesirable consequences, including:

- **Increased inequality:** Supply-side labor market reforms often result in lower incomes for the working class in a nation, as unemployment benefits, labor union power and minimum wages are all reduced. Supply-side tax reforms may also

redistribute the total burden of taxes in a nation away from the higher income earners and more onto the middle and lower income earners

- **Environmental concerns:** Decreased regulation of the private sector may lead firms to find new ways to externalize costs on third parties, often times meaning increased environmental damage
- **Political obstacles:** Increasing spending on education and infrastructure is often times politically difficult due to the fact that the payoff from such investments are often not seen for years or even decades, long after the members of the current government are out of office

3.1 International Trade

The benefits of international trade

- Explain that gains from trade include lower prices for consumers, greater choice for consumers, the ability of producers to benefit from economies of scale, the ability to acquire needed resources, a more efficient allocation of resources, increased competition, and a source of foreign exchange.

Introduction to international trade

The market economic system is based on the premise that through voluntary exchanges between individuals specializing in the production of particular goods and services, both individuals and society as a whole will enjoy a higher level of total output, more efficiency, and higher living standards than would be possible if every individual attempted to live a life of “self-sufficiency”.

It just doesn't make sense in a world of technology and industrialized production for an individual to try to produce everything for himself. Rather, individuals tend to specialize in certain “trades”. I am a teacher. I specialize in teaching economics to young learners. I earn money selling my services to other teachers and to students. I then use that money to buy nearly everything else I need to support myself and my family: food, shelter, clothing, recreation, and so on.

Just as individuals can increase their own well-being through specialization and trade, so can nations as a whole. On the macro level, individual countries will enjoy higher total output, greater efficiency, and higher living standards if they allocate their resources towards the production of goods and services that they are particularly good at producing, just like I allocate my time and energy towards teaching economics.

Through specialization and trade, countries can experience several benefits that they would miss out on if they tried to be “self-sufficient”:

- Lower prices and greater choice for consumers:** When another country can produce a particular good more cheaply, it makes sense to import from that country. In this way, consumers can benefit from lower production costs and lower prices achievable abroad.
- Economies of scale:** When producers in a country specialize in the production of certain types of goods and export their surplus output to other countries, they can produce more output and achieve greater efficiency than if they only produced for the domestic market. Economies of scale (defined in the micro section of this course) refer to the cost advantages enjoyed by firms that produce at a higher level of output.
- The ability to acquire needed resources:** Some countries are more well endowed with natural resources than others. A small island nation like Singapore would be extremely poor if it did not trade with other countries. Singapore is a tiny (30 km

across) island possessing little in terms of natural resources. However, through trade, Singapore has become one of the richest countries in the world, primarily because it is able to obtain resources it otherwise would not possess.

- **More efficient allocation of resources and increased competition:** When countries specialize in production of goods they are particularly good at producing, resource allocation becomes more efficient on a global level. Just as in microeconomics, where increased competition means more efficiency, in international trade the more countries there are competing with one another, the more efficient each country must be with its resources.
- **A source of foreign exchange:** If a country wishes to import goods from other countries, it must also export to those countries. It is through the sale of their own goods that a country earns the foreign exchange (currency) needed to buy other countries' goods.

HL only objectives:

- Explain the theory of absolute advantage.
- Explain, using a diagram, the gains from trade arising from a country's absolute advantage in the production of a good.
- Explain the theory of comparative advantage.
- Describe the sources of comparative advantage, including the differences between countries in factor endowments and the levels of technology.
- Draw a diagram to show comparative advantage.
- Calculate opportunity costs from a set of data in order to identify comparative advantage.
- Draw a diagram to illustrate comparative advantage from a set of data.
- Discuss the real-world relevance and limitations of the theory of comparative advantage, considering factors including the assumptions on which it rests, and the costs and benefits of specialization.

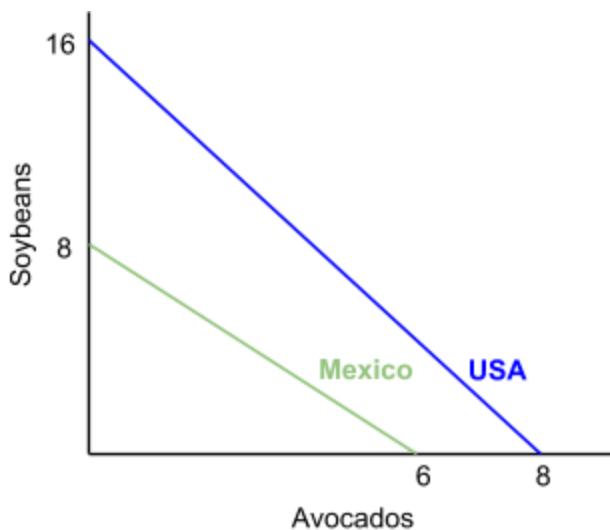
Absolute advantage (HL only)

An individual or a nation has an **absolute advantage** in a good's production when it requires fewer resources to produce the good than the other individual or country.

The table below shows the production possibilities of two countries, the United States and Mexico, for two goods, soybeans and avocados. The values indicate the potential output of each country in each good on a single acre of land.

	soybeans	avocados
Mexico	8 tons	6 bushels
United States	16 tons	8 bushels

The data in the table can also be shown in PPC models for the two countries.



The production possibilities table and the PPC graphs for the USA and Mexico indicate that the United States has an absolute advantage in the production of both avocados and in soybeans, since it can produce more of both goods on a single acre of land than Mexico can.

Comparative advantage (HL only)

An individual or a nation has a **comparative advantage** when it can produce a good or service at a lower opportunity cost than another producer. To determine who has the comparative advantage in avocados and soybeans, we must calculate the opportunity costs of both goods in the USA and in Mexico.

Avocados:

- The USA can produce either 8 bushels of avocados or 16 tons of soybeans on each acre of land. Therefore, the opportunity cost of each bushel of avocados is $16 \div 8 = 2$ **tons of soybeans per bushel of avocados**
- Mexico can produce either 6 bushels of avocados or 8 tons of soybeans on each acre of land. Therefore, the opportunity cost of each bushel of avocados is $8 \div 6 = 1.33$ **tons of soybeans per bushel of avocados**

Avocados can be produced at a lower opportunity cost in Mexico than in the USA. Therefore, Mexico has a comparative advantage in avocados.

Soybeans:

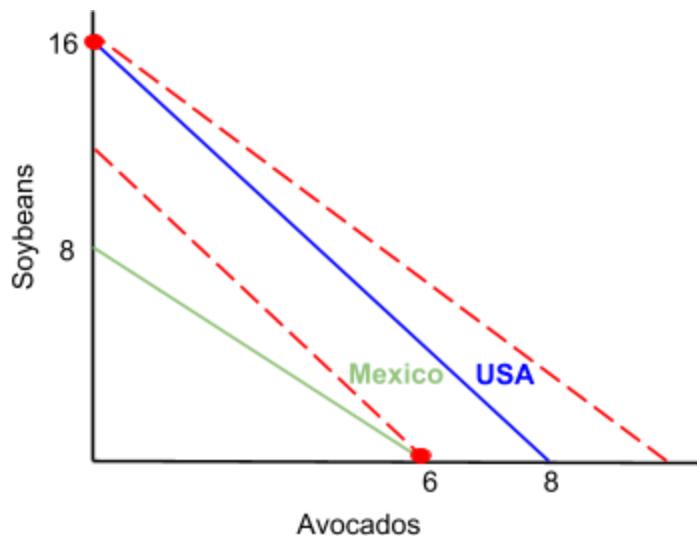
- The USA can produce either 16 tons of soybeans or 8 bushels of avocados on each acre of land. Therefore, the opportunity cost of each ton of soybeans is $8 \div 16 = 0.5$ **bushels of avocados per ton of soybeans**
- Mexico can produce either 8 tons of soybeans or 6 bushels of avocados on each acre of land. Therefore, the opportunity cost of each ton of soybean is $6 \div 8 = 0.75$ **bushels of avocados per ton of soybeans**

Soybeans can be produced at a lower opportunity cost in the USA than in Mexico. Therefore, the USA has a comparative advantage in soybeans.

Specialization based on comparative advantage (HL only)

When individuals or countries specialize in the production of the goods for which they have the comparative advantage, exchange opportunities arise that lead to production possibilities beyond the PPC.

For example, assume that instead of producing both soybeans and avocados domestically, the USA chooses to specialize all its agricultural land in soybean production. Likewise, Mexico allocates all its fertile land towards avocados. The countries can now trade with one another and enjoy a higher level of consumption than would have been possible before trade.



The PPCs above show the outcome when the US and Mexico specialize and trade. The red dots represent each country's actual level of domestic production. The US specializes in soybeans and Mexico in avocados. The red dashed lines represent the possible levels of consumption the two countries can enjoy when they trade with one another. Notice that through specialization and trade the countries enjoy consumption possibilities that are beyond their domestic production possibilities.

Countries and individuals that specialize and trade with others enjoy gains from trade through increased consumption possibilities.

Determining advantageous terms of trade (HL only)

In order to **gain from trade**, an individual or country must be able to acquire a good that they are trading for more cheaply than they would have to pay for it if they produced it on their own.

For example, before trade a single avocado “cost” the US two soybeans. If the US can import avocados for less than two soybeans, then the US gains from trade.

Assume the US and Mexico settle on a “**trading price**” of 1.5 tons of soybeans per bushel of avocados. (which is equal to 0.67 bushels of avocados per ton of soybeans). At this exchange price, the US can import avocados more cheaply than they could produce them for domestically. Likewise, Mexico can import soybeans more cheaply than they can produce them for themselves (remember, one ton of soybeans cost Mexico 0.75 bushels of avocados before trade).

The table below shows the consumption possibilities of the US and Mexico with and without trade (assuming a terms of trade of 1.5 soybeans per avocado).

	soybeans (without trade)	avocados (without trade)	soybeans (with trade)	avocados (with trade)
Mexico	8	6	$6 \times 1.5 = 9$	6
United States	16	8	16	$16 \times 0.67 = 10.7$

The red numbers show that with trade both countries have benefited from a lower opportunity cost for the good they are importing, allowing them to consume a greater total combination of both goods with trade than they would have been able to without trade.

Specialization and trade based on the principle of comparative advantage, either between individuals or between entire nations, will lead to increases in efficiency and a higher level of consumption between those who participate in it.

The gains from trade in supply and demand diagrams (HL and SL)

Free trade allows countries to consume at a level beyond what is possible given their domestic production possibilities. The gains from trade can also be illustrated in supply and demand diagrams.

Assume, for example, that the United States and China engage in trade in two goods: steel and airplanes, and that China has a comparative advantage in steel while the US has a comparative advantage in airplanes.

The US will specialize in manufacturing airplanes and China steel. The two will trade with one another, and both will benefit in the following ways:

In steel:

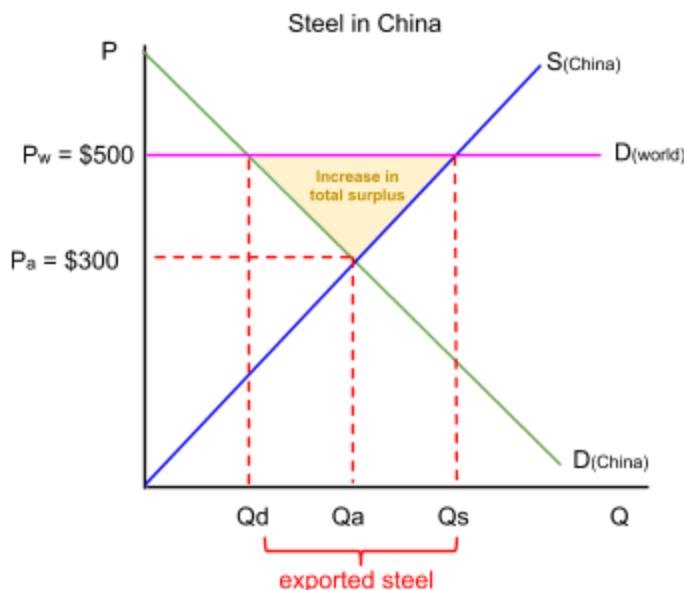
- China will export steel at a price higher than it would be able to sell it for if it were only producing for the domestic market.
- The US will import steel at a price lower than it would be able to buy it for if it were only consuming domestically produced steel.

Airplanes:

- The US will export airplanes at a higher price than it would be able to sell them for if it were only producing for the domestic market.
- China will import airplanes at a lower price than it would be able to if it were only consuming domestically produced airplanes.

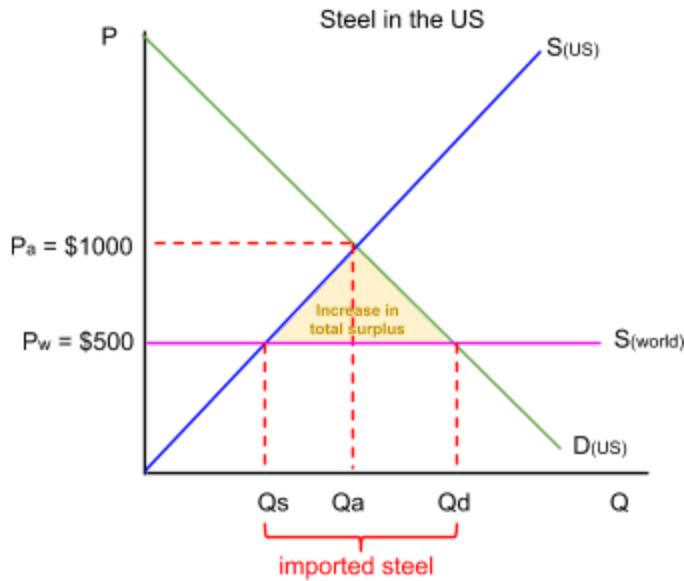
These gains in producer and consumer surplus and total welfare can be shown in supply and demand diagrams. Let's first consider the steel markets in China and the US.

China, which exports steel, benefits from a higher world price than its domestic price, as in the graph below.



Without trade, China would produce Q_a steel at a price of \$300. However, because the US and other countries also demand Chinese steel (D_{world}), China is able to export steel at \$500 and produce Q_s . The higher price and greater quantity benefits Chinese producers, who enjoy greater producer surplus. The effect on overall welfare is illustrated by the yellow triangle.

The gains from trade in steel are also visible in the US steel market.



Without trade, the US would produce Q_a steel at a price of \$1,000. However, because China and other countries also produce steel (S_{world}), the US is able to import steel at \$500 and buy a quantity of Q_s . The lower price and greater quantity benefit American consumers, who enjoy greater consumer surplus. The effect on overall welfare is illustrated by the yellow triangle.

In the steel market US consumers benefit and Chinese producers benefit. In the market for airplanes, it is US producers who enjoy higher prices, greater output, and more producer surplus. Chinese consumers, on the other hand, are able to import planes and enjoy lower air travel costs thanks to the United States' comparative advantage in airplane manufacturing.

The losers from free trade

Of course, in both the steel and airplane examples above, there are stakeholders who are made worse off because of free trade.

- US steel manufacturers produce a smaller quantity (Q_d) and must sell for a lower price (\$500) than they would if the US steel market were closed to trade with China.
- Chinese airplane manufacturers produce fewer planes and must sell for a lower price than they would if the Chinese airplane market were closed to trade with the US.

When powerful firms in industries like steel in the United States and airplanes in China are harmed by free trade, you can bet there is pressure on governments to offer protection to these industries. In the next section of this chapter we'll look at different types of trade protection governments employ to shelter domestic producers from the harmful effects of free trade.

Restrictions on free trade - types of trade protection

- Explain, using a tariff diagram, the effects of imposing a tariff on imported goods on different stakeholders, including domestic producers, foreign producers, consumers

- and the government.
- Explain, using a diagram, the effects of setting a quota on foreign producers on different stakeholders, including domestic producers, foreign producers, consumers and the government.
 - Explain, using a diagram, the effects of giving a subsidy to domestic producers on different stakeholders, including domestic producers, foreign producers, consumers and the government.
 - Describe administrative barriers that may be used as a means of protection.
 - Evaluate the effect of different types of trade protection.

Introduction to protectionism

Despite the gains from trade we have explained and illustrated using both PPCs and supply and demand diagrams, almost every country still chooses to engage in some degree of protectionism.

Protectionism refers to the use of tariffs, quotas, subsidies or administrative measures aimed at making domestic producers more competitive with foreign producers by limiting the quantity of imports into the nation.

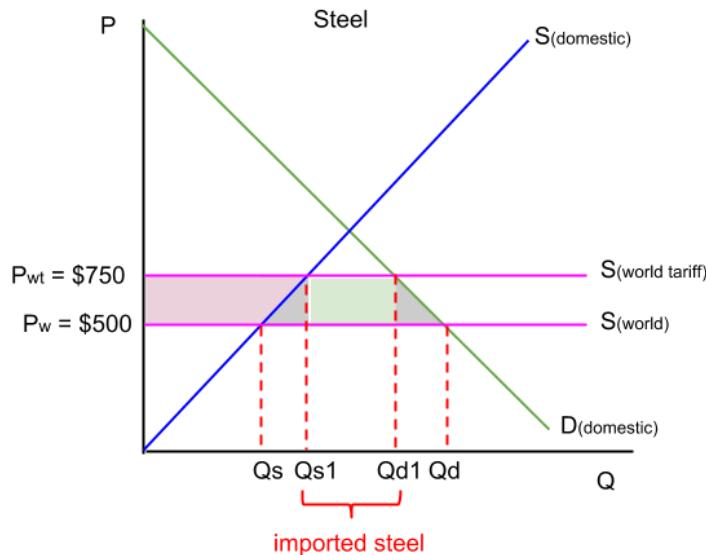
- **Tariffs** are taxes placed on imported goods, services or resources. A tariff increases the cost of imported goods, reducing their supply and causing the price paid by domestic consumers to rise. Therefore, the domestic quantity supplied is greater than it would be without the tariff.
- A **quota** is a physical limit on the quantity of a good, service or resource that may be imported. A quota on a particular good will result in a shortage of imports in the short-run, which drives up the domestic price and leads domestic producers to increase their quantity supplied.
- **Protective subsidies** are payments from the government to domestic producers meant to either increase domestic consumption of their goods or to promote the export of their goods to the rest of the world. The subsidy increases the domestic supply of a good and therefore increases the quantity consumed by domestic consumers.

All forms of protectionism lead to a misallocation of society's resources and ultimately reduce total welfare. However, there are several arguments for protectionism that must be evaluated.

Tariffs

In order to protect domestic producer from lower cost imports, governments sometimes employ tariffs. A **tariff** is a tax on imports.

Assume the United States government wished to shelter US steel producers from cheaper Chinese imported steel. A steel tariff would raise the world price in the domestic market and reduce the world supply, leading to an increase in domestic quantity supplied.



The \$250 tariff on each ton of imported steels has caused:

- The domestic quantity supplied to increase from Q_s to Q_{s1}
- The domestic quantity demanded to fall from Q_d to Q_{d1}
- The quantity of imported steel to decrease to $Q_{d1} - Q_s$
- An increase in domestic producer surplus equal to the pink area
- An increase in government tax revenues equal to the green area
- A deadweight loss equal to the gray areas.

The tariff is inefficient because the decrease in consumer surplus it causes is greater than the increase in producers surplus and the increase in government revenue.

Tariffs are controversial forms of protectionism for many reasons.

- They are usually passed onto consumers of the protected good, causing domestic prices to rise in certain industries
- They harm foreign producers and may be met with retaliation by foreign governments
- They protect domestic industries that may not have a comparative advantage, thus may make an economy less efficient over time

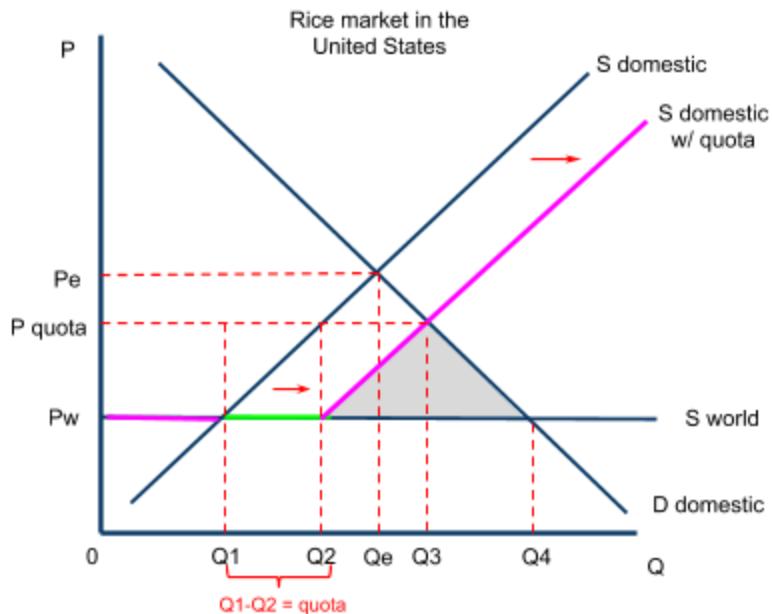
Quotas

Another protectionist policy is a **quota**. Quotas are physical limits on how much of a good can be produced or imported. A protectionist quota limits the quantity of imports of a particular commodity or good into a country, thereby protecting domestic producers from foreign competition.

Quotas create a shortage in the short term, causing price to rise and domestic producers to fill the gap. In the end, consumers pay a higher price and there is less output and less consumer surplus. Producers benefit because they control a larger share of the domestic market for their good or commodity. Foreign firms tend to prefer quotas over tariffs,

because instead of the higher price consumers pay going to the government, importers also sell for the higher price, albeit at a lower quantity than under free trade.

The graph below shows the effect of a quota of Q_1-Q_2 units of rice imported into the United States



Impact of the quota on market equilibrium:

- Before the quota Q_1-Q_4 units of rice would have been imported at a world price of P_w . Q_1 units would have been produced by US growers
- After a quota of Q_1-Q_2 units is imposed, the world supply is limited to the green line
- There is a short-term shortage of rice os Q_2-Q_4 units
- The shortage causes the price to rise, leading to an increase in domestic quantity supplied
- Price increases to P_{quota} , and the domestic quantity supplied is now $0-Q_1$ and Q_2-Q_3

Impact of the quota on total surplus

- Domestic producers benefit because they sell more rice under the quota at a higher price
- Foreign producers suffer because they sell less rice under the quota (albeit at a higher price)
- Domestic consumers are worse off as the price is higher and the quantity is lower; there is less consumer surplus
- Overall there is a deadweight loss equal to the gray triangle; the loss of consumer surplus is greater than the increase in producer surplus

As with all forms of protectionism, quotas lead to a reallocation of welfare from one

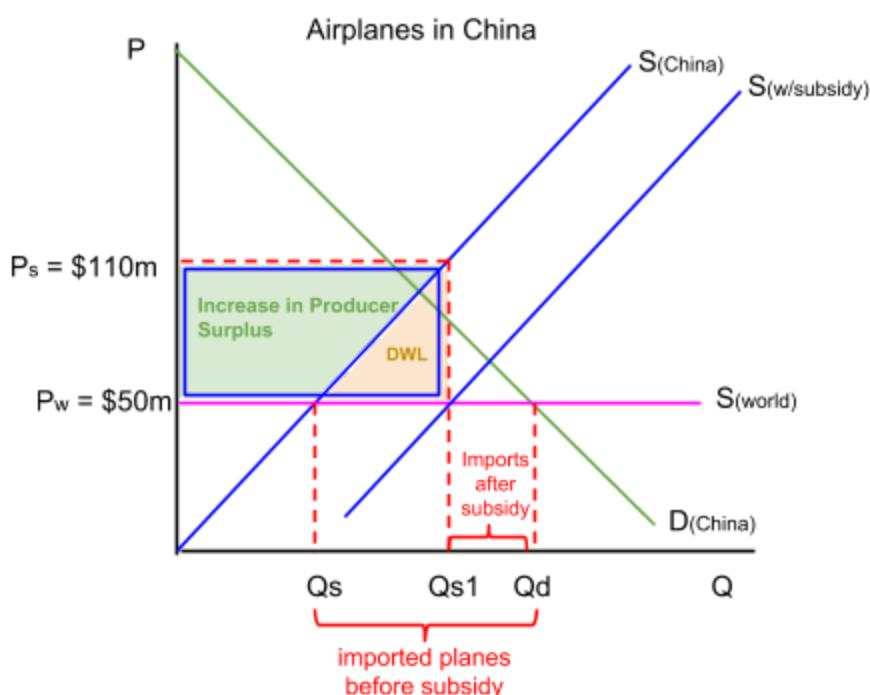
stakeholder to another; in this case domestic consumers and foreign producers are worse off while domestic producers are better off. Also, as with other forms of protectionism, a quota results in a deadweight loss and an overall misallocation of resources both in the domestic economy and abroad.

Protectionist subsidies

A third form of protectionism involves government subsidizing domestic producers with the aim of reducing their costs to a level that makes them more competitive with cheaper foreign producers.

Assume the Chinese government decides that it wants more of the airplanes bought by Chinese airlines to be produced in China. To reduce imports of airplanes from the US, China offers massive subsidies to Chinese firms producing airplanes in China.

The graph below shows the effect of a subsidy for Chinese plane manufacturers assuming China has been importing airplanes at a lower world price from the US and other countries.



In the graph above:

- China has granted Chinese plane producers a \$60 million subsidy for each plane produced.
- Chinese airplane buyers still pay the world price of \$50 million, but producers receive \$60 million on top of this, earning producers \$110 million per plane sold.
- Without the subsidy, China would import $Q_s - Q_d$ airplanes. Following the subsidy, China imports only $Q_{s1} - Q_d$ airplanes.
- Chinese consumers are no worse off, while producer surplus increases by the green area.

- The blue rectangle represents the cost of the subsidy to the Chinese government ($Qs1 \times \$60m$).
- The subsidy's cost (blue rectangle) is greater than the subsidy's benefit (green area) by the amount represented by the pink triangle (the deadweight loss).

Unlike tariffs, protectionist subsidies make domestic consumers no worse off, since the price paid does not change. However, because subsidies cost government money, there is a deadweight loss because the resulting increase in producer surplus is less than the cost to the government.

Administrative barriers to trade

Not all protectionism involved intervention in the markets for goods through subsidies, quotas, or tariffs. “Non-tariff” and other “administrative” barriers to trade include:

- **Licenses:** Governments may require firms importing their goods into a country to acquire a limited number of licenses for import. Such administrative controls limit the quantity of imports coming into a country by limiting the number of firms allowed to import.
- **Quality or environmental standards:** By setting standards of product quality, safety standards, or “green” manufacturing standards, a country can limit the quantity of goods coming in from producers in lower-cost countries, where manufacturing standards may be lower.
- **Foreign exchange intervention:** Manipulating its exchange rate on the foreign exchange market can give a country an advantage in international trade. “Devaluation” of the currency (making it weaker against other currencies) makes a country’s exports cheaper in foreign countries, and makes imports more expensive at home, limiting imports and promoting exports. Foreign exchange markets are explored later in this chapter.
- **Voluntary export restraints:** In some cases a government may request that the government of a trading partner voluntarily limit the quantity of goods it exports. To encourage voluntary export restraints, tariffs may be threatened.

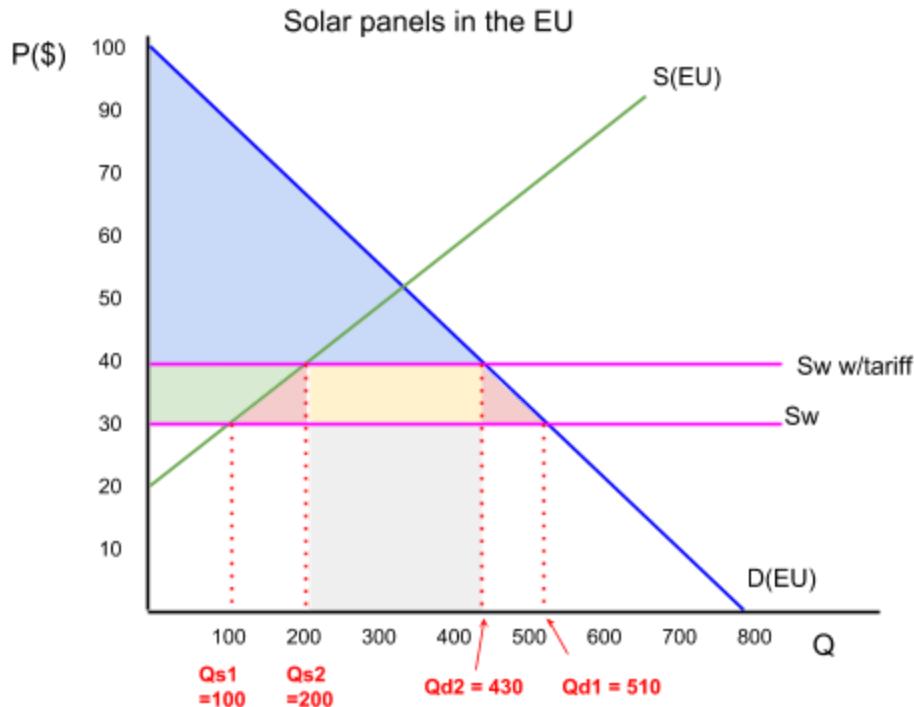
HL only objectives

- Calculate from diagrams the effects of imposing a tariff on imported goods on different stakeholders, including domestic producers, foreign producers, consumers and the government.
- Calculate from diagrams the effects of setting a quota on foreign producers on different stakeholders, including domestic producers, foreign producers, consumers and the government.
- Calculate from diagrams the effects of giving a subsidy to domestic producers on different stakeholders, including domestic producers, foreign producers, consumers and the government.

The effects of protectionist measures can be calculated from a diagram

Tariffs

The graph below shows the effect on price and quantity a \$10 tariff imposed on imported solar panels in the European Union.



Using values on the graph, we can calculate different effects of the tariff. For example:

- The increase in producer surplus caused by the tariff (green area):

$$= (\$10 \times 100) + \frac{\$10 \times 100}{2} = \$1,500$$
- The government revenue generated by the tariff (yellow area):

$$= \$10 \times 230 = \$2,300$$
- The deadweight loss of the tariff (red areas):

$$= \frac{\$10 \times 100}{2} + \frac{\$10 \times 80}{2} = \$500 + \$400 = \$900$$
- Foreign solar panel producer revenues after the tariff (gray area):

$$\$30 \times 230 = \$6,900$$
- Consumer surplus after the tariff (blue area):

$$\frac{(\$100 - \$40) \times 430}{2} = \frac{\$60 \times 430}{2} = \frac{\$25,800}{2} = \$12,900$$

Given price and quantity values on a protectionist graph, the impact of a tariff, quota, or subsidy can be calculated by determining the areas consumers surplus, producer surplus, deadweight loss, producer revenues, government revenue, and so on.

Arguments for and against trade protection

- Discuss the arguments in favour of trade protection, including the protection of domestic jobs, national security, protection of infant industries, the maintenance of health, safety and environmental standards, anti-dumping and unfair competition, a means of overcoming a balance of payments deficit and a source of government

- revenue.
- Discuss the arguments against trade protection, including a misallocation of resources, the danger of retaliation and “trade wars”, the potential for corruption, increased costs of production due to lack of competition, higher prices for domestic consumers, increased costs of imported factors of production and reduced export competitiveness.

Protectionism always leads to a loss of total surplus (or deadweight loss). So why do countries still practice it? Here are some of the arguments for and against protectionism.

Arguments for protectionism:
<ul style="list-style-type: none"> Protection of domestic employment: firms in protected industries produce more and employ more than they would if forced to compete on a level playing field with cheaper foreign producers. Protection of infant industries: Young, emerging industries are able to grow under government protection until they can compete with foreign producers. To prevent dumping: If foreign firms are selling their products in the domestic market at a price lower than their costs of production, protectionism can actually level the playing field, preventing domestic producers from losing due to this unfair trade practice. To enforce product standards: Protectionism can be used to protect domestic consumers from unsafe or harmful imports. To raise revenue: Tariffs created government revenue, which could help balance a government's budget and provide needed funds for public goods and services. To protect strategic industries: Defense, energy, food and other industries essential for national security may warrant protecting in case conflict or other issues arise that may cut off global supply chains. Reducing imbalances in the current account: A country with a large trade deficit may impose protectionism as a means of reducing the imbalance and avoiding the negative effects associated with trade deficits (more on this later in the chapter)
Arguments against protectionism
<ul style="list-style-type: none"> Leads to a misallocation of resources: Too much of the protected good will be produced domestically, not enough by relatively efficient foreign producers Could lead to a trade war: If trading partners retaliate with their own protections, even worse resource allocation will result. Higher priced imports: In the case of quotas and tariffs consumers suffer from higher prices and some producers will have to pay more for imported raw materials. Reduced competitiveness: Industries sheltered by protectionism will become less and less competitive over time, requiring even more protection and a greater loss of welfare.

Based on all our analysis, some broad conclusions can be made about most forms of protectionism. Ultimately, protectionism creates some winners and losers, but the cost to the losers exceeds the benefits to the winners. Protectionism...

- Benefits: Domestic producers may benefit b/c they receive a higher price for their output. The federal government may gain through revenue from tariffs.
- Harms: Consumers are harmed because they pay higher prices for goods produced by the protected industry. Foreign producers are hurt because they are not able to sell their as much of their output as they would be able to otherwise, so their profits are reduced.

Most economists oppose protectionism. In most cases, the costs of protectionism exceed the benefits. Consumers are hurt by the higher prices they pay, while producers often benefit less. Also, industries employ large amounts of economic resources in “**rent-seeking**”, which refers to the practice of lobbying government to erect barriers to trade in favor of particular stakeholders. In most cases, protectionism results in deadweight loss for society, meaning economic inefficiency.

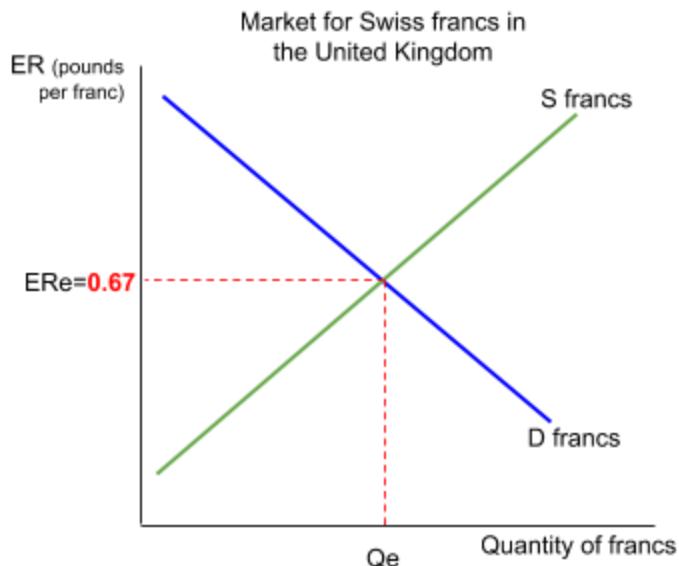
3.2 Exchange Rates

Freely floating exchange rates

- Explain that the value of an exchange rate in a floating system is determined by the demand for, and supply of, a currency.
- Draw a diagram to show determination of exchange rates in a floating exchange rate system.

Introduction to the foreign exchange (forex) market

A currency's exchange rate is determined in the **foreign exchange market** for that currency (also called a forex market). For example, consider the market for Swiss francs in the United Kingdom. The supply and demand graph below represents the Swiss franc forex market in the UK.



Let's examine each of the elements of the forex market for Swiss francs in the UK

- Title:** It's always important to clearly indicate which currency in which country a forex market graph represents. In this case, it is the market FOR Swiss francs IN the United Kingdom. Always give your forex market graphs a clear title!
- Vertical axis label:** All that is required on the vertical axis of a forex graph is “Exchange Rate” or “ER”. However, you may also choose to identify the units represented by the values on the axis. In this case, the units are “pounds per franc.” In other words, it’s how much a single Swiss franc costs in terms of British pounds.
- Horizontal axis label:** Be sure to label your quantity axis with the name of the currency the forex graph represents. Here we labelled it “quantity of francs.”
- Supply and demand curve labels:** It's also a good idea to identify which currency

is being supplied and demanded in this graph. The green line represents the supply of Swiss francs in the UK and the blue line represents the demand for Swiss francs in the UK.

- **Equilibrium exchange rate:** Notice that rather than using “Pe” for the equilibrium, we have used “ERe”, short for “equilibrium exchange rate”. Values are usually not necessary when drawing forex graphs, unless they are known or explicitly required by whatever problem you are solving.

Who demands a currency?

The demand for a currency in the forex market represents specific individuals and groups of individuals in a country. For example, the demand for Swiss francs in the UK represents:

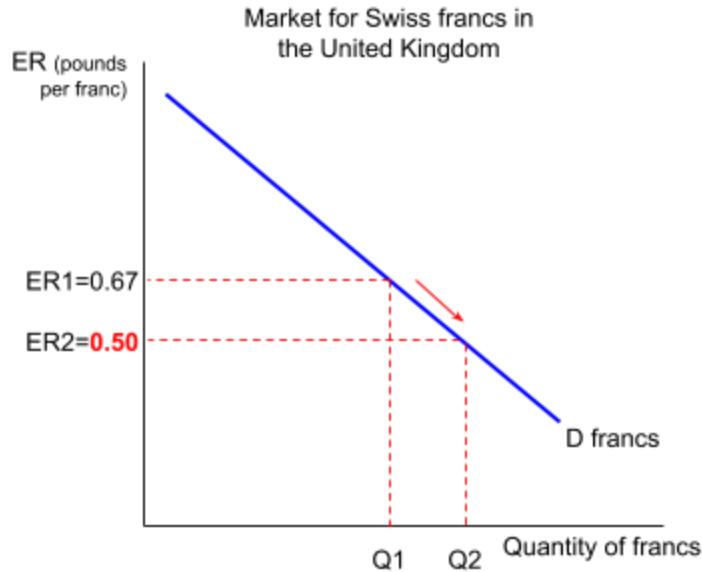
- **Consumers:** British consumers who wish to buy Swiss goods or travel to Switzerland demand Swiss francs with which to buy Swiss goods and services.
- **Firms:** British firms that wish to import technology, raw materials, or other inputs from Switzerland demand Swiss francs with which to buy Swiss inputs. British firms that wish to operate offices or build manufacturing facilities in Switzerland also require francs with which they can acquire capital in Switzerland.
- **Investors:** British investors who wish to invest in Swiss assets (government bonds, stocks in Swiss companies, Swiss bank accounts) demand francs with which they can acquire Swiss financial assets.
- **Central banks:** The British central bank (the Bank of England) demands Swiss francs to hold onto in case it wishes to someday intervene in the forex market to strengthen or weaken the British pound against the franc.

The law of demand in a forex market

You'll recall from chapter one in this text that the **law of demand** says there is an inverse relationship between a good's price and the quantity demand. The downward sloping demand curve for a good reflects this relationship.

You'll notice, of course, that the demand curve for a currency is also downward sloping. This is just a reflection of the law of demand as it applies to international trade.

Consider the Swiss franc. At present the exchange rate is 0.67 pounds per franc, at which Q_e francs are demanded. How would a depreciation of the franc affect the quantity demanded, and why would it affect it that way?



When the franc's exchange rate falls from 0.67 pounds to 0.50 pounds, the quantity demanded increases from Q1 to Q2. Reasons for this are:

- Swiss goods have gotten cheaper when the franc depreciates, so the quantity demanded of Swiss imports to the UK increases. More imports require more francs.
- Swiss assets (bonds, stocks, etc.) have gotten cheaper for British investors, so the quantity demanded of Swiss assets in the UK increases. More investment in Switzerland requires more francs.

The law of demand applies in forex markets just as it does in product markets. When a currency gets cheaper, more of it is demanded in the country in which it has depreciated.

Who supplies a currency?

So we know that British consumers, firms, and investors demand Swiss francs, but who supplies Swiss francs to these people? Well, the supply of a currency is made possible by foreigners whose currency we're considering. In the UK, Swiss francs are supplied by people from Switzerland, specifically:

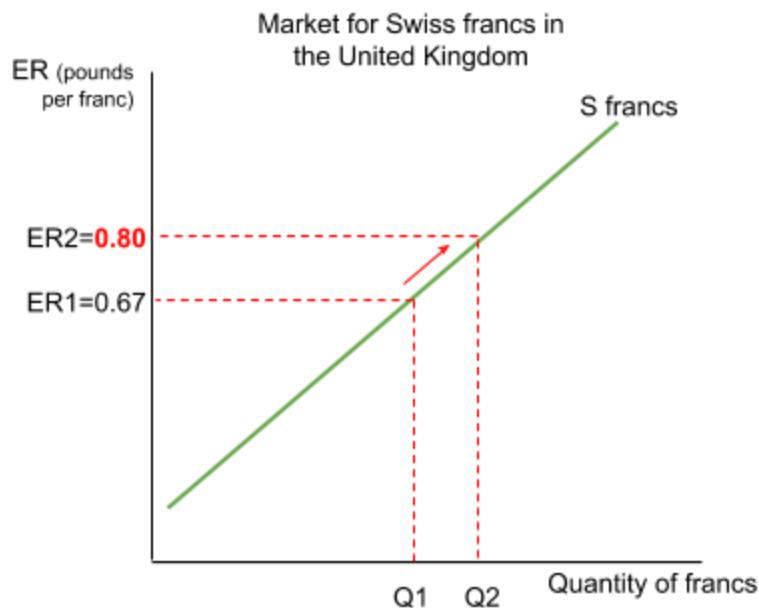
- **Consumers:** Foreign consumers who demand imports must supply their own currency to the forex market in exchange for the currency of the country whose goods they want to buy. Swiss consumers supply francs to the UK, which they exchange for British pounds to buy British goods.
- **Firms:** Foreign firms that demand raw materials, capital, or other factors of production from foreign countries must supply their own currency to the forex market in exchange for the currency of the country whose imports they want to buy. Swiss firms supply francs to the UK, which they exchange for British pounds.
- **Investors:** Foreign investors who wish to invest in a country's assets must supply their currency to the forex market, where they exchange it for the domestic country's currency to invest in domestic assets. Swiss investors supply francs to the UK, which they exchange for British pounds in order to invest in British bonds, stocks, and

- other financial assets.
- **Central banks:** Foreign central banks that demand foreign currency for their own reserves must supply their currency to the forex markets in exchange for the foreign currency. The Swiss central bank (the Swiss National Bank), supplies francs to the UK, which they exchange for British pounds in order to hold pounds in its foreign exchange reserves.

The law of supply in a forex market

Recall from chapter one that the **law of supply** says that there is a direct relationship between a good's price and the quantity supplied. The same principle applies to foreign exchange.

Consider the forex market for the Swiss franc gain. At the current exchange rate of 0.67 pounds, Q_e Swiss francs are supplied by Swiss households, firms, and investors. How would an appreciation of the franc affect the quantity of francs supplied to the forex market?



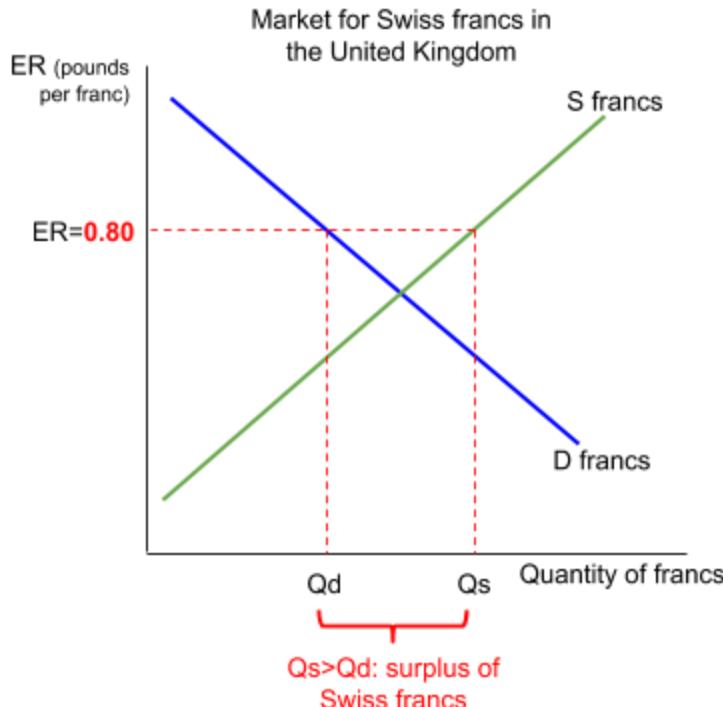
When the franc's exchange rate rises from 0.67 pounds to 0.80 pounds, the quantity supplied increases from Q₁ to Q₂. Reasons for this are:

- British goods have gotten cheaper to Swiss consumers as the value of the franc has increased in the UK, so the Swiss are willing to supply more francs at a higher exchange rate (so they can buy more British goods!)
- British assets (bonds, stocks, etc.) have gotten cheaper to Swiss investors as the franc has appreciated, so the Swiss are willing to supply more francs at a higher exchange rate (so they can invest in more British assets!)

Equilibrium and disequilibrium in the forex market

A forex market is in **equilibrium** when the exchange rate is at the level where the quantity demanded by domestic stakeholders is equal to the quantity supplied by foreign stakeholders.

In the forex market for Swiss francs in the UK, the equilibrium exchange rate was 0.67. But what happens if an exchange rate other than this prevails? Assume, for example, that sellers of Swiss francs in the UK attempt to sell francs for 0.80 British pounds per franc. How does this affect the market?



At 0.80 pounds per franc the forex market is in disequilibrium. There is a surplus of Swiss francs supplied on the market.

- Swiss consumers, firms, and investors are willing to supply more francs in the UK than British consumers, firms, and investors demand.
- The exchange rate must fall in order to restore equilibrium.
- As the exchange rate falls the Swiss will be willing to supply less francs and the British will demand more francs, until the market is back at equilibrium at 0.67 pounds per franc.

An exchange rate below equilibrium will create a shortage of a currency and the story will be reversed:

- Swiss consumers, firms, and investors would be willing to supply fewer francs in the UK than British consumers, firms, and investors would demand (remember, a cheap franc means cheap Swiss goods and assets).
- The exchange rate would have to rise in order to restore equilibrium.
- As the exchange rate rises the Swiss would be willing to supply more francs (as British goods and assets become more attractive) and the British would demand

fewer francs (as Swiss goods and assets become more expensive), until the market is back at equilibrium at 0.67 pounds per franc.

Just like in product markets, disequilibrium occurs when the quantities demanded and supplied are not equal to each other. Market forces drive exchange rates towards equilibrium, at which point domestic stakeholders demand the exact amount of a currency that foreign stakeholders are willing to supply.

HL only objectives

- Calculate the value of one currency in terms of another currency.
- Calculate the exchange rate for linear demand and supply functions.
- Plot demand and supply curves for a currency from linear functions and identify the equilibrium exchange rate.
- Using exchange rates, calculate the price of a good in different currencies.

Once you know the value of one currency expressed in terms of another, we can easily calculate the value of the other currency expressed in terms of the original. Study the table below.

1 USD	Euro	British Pound	Indian Rupee	Australian \$	Canadian \$	South African rand	New Zealand \$	Japanese Yen	Chinese yuan
1 US \$ =	0.81	0.64	55.51	0.95	1	8.23	1.23	78.13	6.36
Inverse:	1.23	1.56	0.02	1.05	1	0.12	0.81	0.013	0.16

The first row of the table tells us how much one dollar ‘costs’ in each of the foreign currencies. In other words, it’s the dollar exchange rate in Europe, Britain, India, and so on. The second row tells us how much the foreign currency costs to Americans. For example,

- One euro’s worth of goods from Germany ‘costs’ Americans \$1.23. 100 euros of output would cost \$123.
- One rand worth of South African output would only cost an American \$0.12. But 100 rand of output would cost \$12.
- The value of one currency is always the inverse of the other currency’s value

Calculating prices using exchange rates

With knowledge of exchange rates, we can easily calculate how much a good produced abroad in one currency will cost a foreign consumer who is spending another currency. The table below shows the dollar’s exchange rate in terms of four currencies in the first row and the price of a \$1,000 product to a consumer from each of the four countries in the second row.

1 USD	Euro	British Pound	Indian Rupee	Australian \$
1 US \$ =	0.81 euro	0.64 pound	55.51 rupees	0.95 AU\$
Price of a \$1000 American product in each currency	= 1000×0.81 =810€	= 1000×0.64 =640£	1000×55.51 =5,551r	1000×0.95 =950AU\$

To calculate the price of a good in different currencies, we must multiply the price of the good in US dollars by the exchange rate of the dollar in each country.

Exchange rates and linear demand and supply equations

As with goods in product markets, the demand for and supply of a currency in a foreign exchange market can be expressed using linear equations. From these, we can determine the equilibrium exchange rate and the quantities of the currency that will be traded in a country. Assume the demand and supply for Euros in Switzerland is represented by the equation

$$Q_d = 100 - 50E$$

and supply as:

$$Q_s = 10 + 30E$$

The equations express how many millions of euros would be demanded and supplied at different exchange rates for the euro. “E” is the exchange rates in of euros in Swiss francs.

For every 1 franc increase in the price of a euro, 50 million fewer euros will be demanded by Swiss households. This is because European goods are getting more expensive and thus Swiss demand less of the European currency.

For every 1 franc increase in the value of a euro, Europeans will supply 30 million more euros to the Swiss market. This is because Swiss goods are becoming cheaper to European households, so Europeans supply more euros in order to buy Swiss goods.

With the supply and demand for euros in Switzerland known, we can calculate the equilibrium exchange rate of the euro in Switzerland by setting demand and supply equal to one another and solving for “E”.

$$100 - 50E = 10 + 30E$$

$$90 = 80E$$

$$E=1.125$$

One euro costs a Swiss household 1.125 Swiss francs. With this exchange rate, we know that:

- A car that costs 20,000€ will cost a Swiss consumer 22,500 francs.
- A hotel room in Berlin that costs 100€ per night will cost a Swiss guest 125 francs.

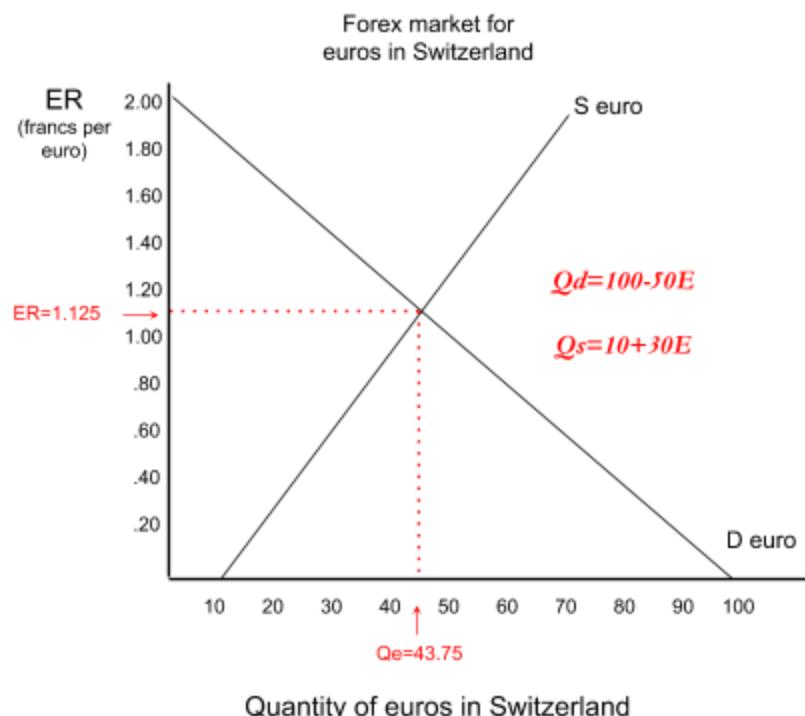
If the euro appreciates, *ceteris paribus*,

- Swiss households will demand fewer euros (50 million less for every 1 franc increase in its value). This is because Swiss consumers will demand fewer European goods and services.
- European households will supply more euros to Switzerland (30 million more for every 1 franc increase in value). This is because European consumers will demand more Swiss goods and services, and therefore supply more euros.

Graphing a currency's demand and supply from equations

As with the demand and supply for a good or service, a forex market can be illustrated using the equations for the currency's demand and supply.

In the graph below, the demand and supply of euros in Switzerland is illustrated.



Notice from the graph:

- The “p-intercept” of demand is found by setting the quantity to zero and solving for the “E” in the demand equation.
- The “q-intercept” of demand is the ‘a’ variable in the demand equation (100).

- The q-intercept of supply is the ‘c’ variable in the supply equation (10).
- The equilibrium quantity of euros in Switzerland is found by plugging the equilibrium exchange rate (1.125) into either equation and solving for “Q”

$$Q_d = 100 - 50(1.125)$$

$$Q_d = 43.75$$

$$Q_s = 10 + 30(1.125)$$

$$Q_s = 43.75$$

Causes of changes in the exchange rate

- Explain the factors that lead to changes in currency demand and supply, including foreign demand for a country’s exports, domestic demand for imports, relative interest rates, relative inflation rates, investment from overseas in a country’s firms (foreign direct investment and portfolio investment) and speculation.
- Distinguish between a depreciation of the currency and an appreciation of the currency.
- Draw diagrams to show changes in the demand for, and supply of, a currency.

HL objective:

- Calculate the changes in the value of a currency from a set of data.

We know that exchange rates are determined by the demand for and supply of currencies. Next we will consider the factors that can cause a change in a currency’s exchange rate?

First, some terminology:

- **Appreciation:** When a currency’s value increases on the forex market relative to another currency do to an increase in demand or a decrease in supply, the currency appreciates.
- **Depreciation:** When a currency’s value decreases on the forex market relative to another currency due to a decrease in demand or an increase in supply, the currency depreciates.

The table below outlines the factors that can cause a shift in the demand for a currency (and the supply of the other currency) in forex markets.

The Determinants of Exchange Rates	
Tastes and preferences	As a country’s exports become more popular among international consumers, demand for its currency will increase and supply of other countries currency in its forex market will increase.

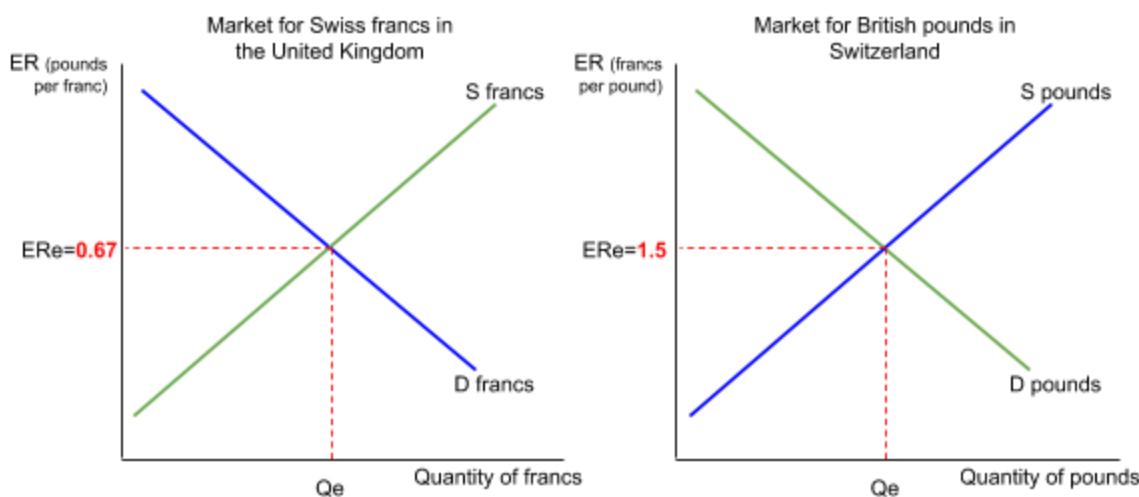
Relative interest rates	There is a direct relationship between the interest rates in a country and the value of its currency. At higher interest rates, foreigners will demand more financial assets from the country, and therefore more of the currency.
Relative price levels (or inflation rates)	If a country's inflation rate is high relative to its trading partners, demand for the country's exports will fall and demand for its currency will fall. If inflation is lower at home than abroad, foreigners will demand more of its exports and its currency.
Speculation	If international financial investors expect a country's currency to appreciate in the future, demand for it will rise today. If a currency is expected to depreciate demand for it will decrease today. Speculation is simply betting on the future value of an asset or currency.
Relative Incomes (or growth rates)	As incomes rise abroad, foreigners will demand more of a country's currency. If foreign incomes fall, there will be less demand for a country's exports and its currency. If domestic incomes rise, ceteris paribus, demand for foreign currencies will rise and supply of the foreign currency will increase abroad, as households wish to buy more imports

To demonstrate, let's consider the market for British pounds in Switzerland. Recall that the value of one currency is the inverse of the other currency's exchange rate. Since francs cost 0.67 pounds, the exchange rate of the pound in Switzerland is the reciprocal of this:

$$1 \text{ pound} = \frac{1}{0.67} \text{ francs}$$

$$1 \text{ pound} = 1.5 \text{ francs}$$

We can graph the market for pounds in Switzerland and the market for francs in the UK side by side:



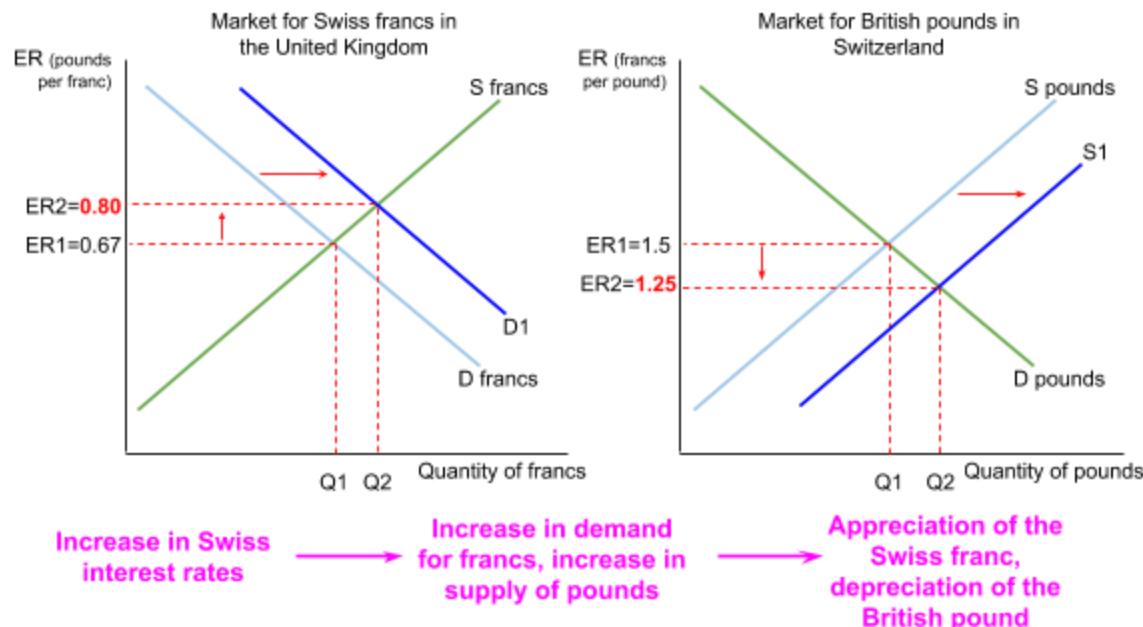
Notice that the demand for francs and the supply of pounds are both **blue**. This is because

both represent the same stakeholders: British consumers, firms, and investors.

The supply of francs and the demand for pounds are both **green**. This is because both represent the same stakeholders: Swiss consumers, firms, and investors.

Any change that causes a shift in the demand for one currency will cause a shift in the supply of the other currency. Consider, for example, the effect of an increase in interest rates in Switzerland:

- Higher interest rates in Switzerland lead British investors to demand more Swiss assets (which now earn a higher return). Demand for Swiss francs increases in the market for francs in the UK.
- In order to acquire the francs they need to invest in Switzerland, British investors must convert their pounds to francs, so the supply of British pounds increases in the market for pounds in Switzerland.
- The demand for francs and the supply of pounds both represent British stakeholders. Both blue lines shift following an increase in interest rates in Switzerland.



Higher returns on Swiss assets have led British investors to demand more francs and supply more pounds. As a result, the franc has appreciated and the pound has depreciated.

Notice that the new exchange rates are reciprocals of one another:

$$1 \text{ pound} = 1.25 \text{ francs}$$

$$1 \text{ franc} = \frac{1}{1.25} \text{ pounds} = 0.8 \text{ pounds}$$

Fiscal policy's effect on the exchange rate

As we learned in an earlier chapter, **expansionary and contractionary fiscal policies** lead to changes in a country's national output, income level, and the price level. Indirectly, changes in fiscal policy can also affect the exchange rate and a country's net exports.

Expansionary fiscal policy can affect the exchange rate in three ways:

- Increased aggregate demand **drives up the price level** and can make a country's exports less attractive to foreign consumers. Consequently, demand for the country's goods falls and demand for the currency falls, **causing a depreciation** of the currency.
- Increased AD **increases incomes** and employment, which leads to more demand for imports among the country's households. More import demand increase the currency's supply on forex markets, **causing a depreciation** of the currency.
- The effect of expansionary fiscal policy on interest rates was explained in the section on **crowding out**. Deficit spending drives up domestic interest rates, which attracts international financial investors to the country's assets, such as government bonds. The increased demand for the country's assets **causes an appreciation** of the currency on forex markets.

The different effects of expansionary fiscal policy on the exchange rate work against one another, but in all cases, the effect on net exports is negative. As a government pursues fiscal stimulus, the impact on international trade flows counteract the intended effect of the stimulus. Higher prices, higher incomes, and higher interest rates all lead to a **negative net export effect** of fiscal policy. When the government wishes to increase AD, net exports tend to fall.

Contractionary fiscal policies have the opposite effects on the exchange rate as those described above. When the government wishes to decrease AD, net exports tend to rise.

- Higher taxes or reduced government spending causes the price level to fall, increasing demand for the country's exports and its currency, **causing it to appreciate**.
- Lower incomes cause a fall in demand for imports, reducing supply of the currency on forex markets **causing it to appreciate**.
- Finally, smaller budget deficits reduce domestic interest rates and **cause the currency to depreciate** as international demand for financial assets falls.

When the government intends to reduce aggregate demand through fiscal policies, the impact on net exports works in the opposite direction, working against the desired effect of the contractionary policy.

Monetary policy's effect on the exchange rate

While fiscal policy creates a negative net export effect (meaning the change in net exports of a certain fiscal policy works against the policy's intended effect), the use of monetary policies creates a **positive net export effect**.

Consider a central bank's expansion of the money supply during a recession:

- Lower interest rates lead to a fall in savings and an increase in interest sensitive consumption and investment.
- Lower interest rates reduce international demand for the country's financial assets and cause demand for the currency on forex markets to fall, **causing a depreciation of the currency.**
- A weaker currency leads to an increase in demand for the country's goods among foreign consumers, and a fall in demand for imports, causing an **increase in net exports.**
- Aggregate demand increases due to increased consumption, investment, and net exports.

The positive net export effect of monetary policy means that the impact on net exports of an expansionary or contractionary central bank policy reinforces the policy's intended effect on aggregate demand.

Changes in demand or supply in linear equations (HL only)

The supply or demand for a currency may change if one of the determinants of exchange rates changes. This will result in a new equation and a new equilibrium exchange rate.

Assume that there is a financial crisis in Europe and Swiss investors wish to hold less European assets. As a result, the demand for euros decreases in Switzerland. At the same time, European investors demand more Swiss assets and therefore the supply of euros in Switzerland increases. The new equations for demand and supply are:

$$Q_d = 80 - 50E \text{ and... } Q_s = 15 + 40E$$

The new equilibrium exchange rate for the Euro can be calculated

$$80 - 50E = 15 + 40E$$

$$65 = 90E$$

$$E = 0.72$$

The equilibrium quantity can be found by plugging the new exchange rate into the demand and supply equations:

$$Q_d = 80 - 50(0.72)$$

$$Q_d = 43.9$$

and

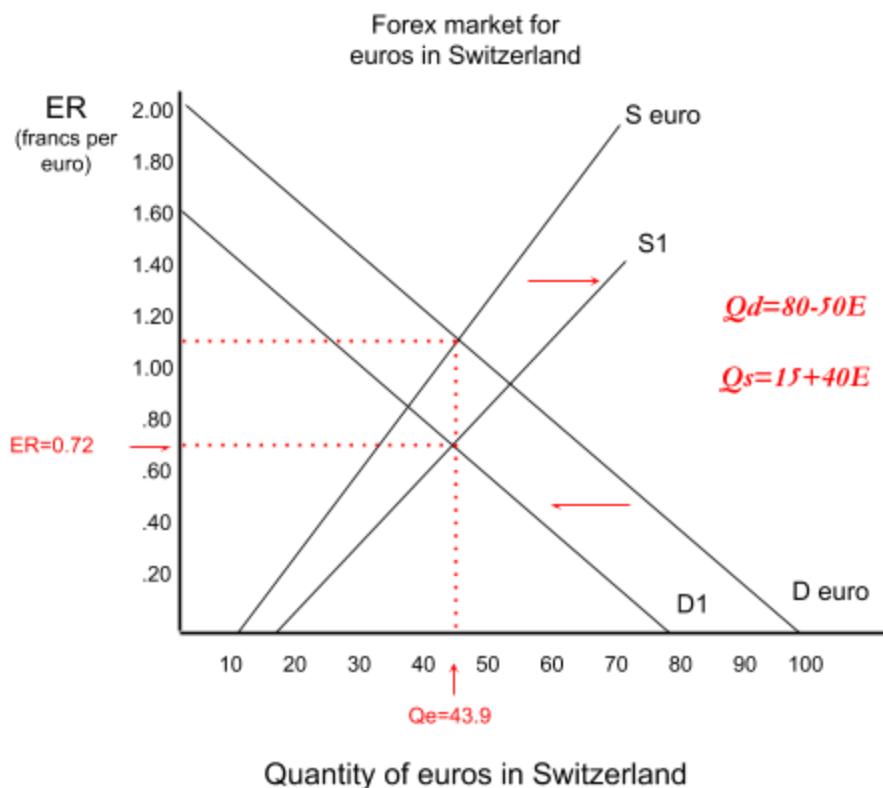
$$Q_s = 15 + 40(0.72)$$

$$Q_s = 43.9$$

The euro has depreciated against the Swiss franc.

- Because demand for it has fallen while its supply increased, the euro is much less scarce in Switzerland
- European goods can be bought much more cheaply as a result (only 0.8125 CHF must be given for a euro's worth of output instead of 1.125 CHF)

These changes in demand and supply can be illustrated in our forex market diagram:



Government intervention in forex markets - fixed exchange rates

- Describe a fixed exchange rate system involving commitment to a single fixed rate.
- Distinguish between a devaluation of a currency and a revaluation of a currency.
- Explain, using a diagram, how a fixed exchange rate is maintained.

We have just examined how exchange rates are determined in a **floating exchange rate system**, in which a currency's value is determined by free market demand and supply of the currency from investors, foreign consumers, central banks, and other stakeholders.

Floating exchange rate systems are common in many forex markets. By allowing its currency to float a country signals to other countries that it embraces free trade and will not

manipulate its exchange rate in order to benefit certain stakeholders.

However, not all currencies' values are determined in a floating exchange rate system. Some countries choose to fix or manage their exchange rates in the forex market.

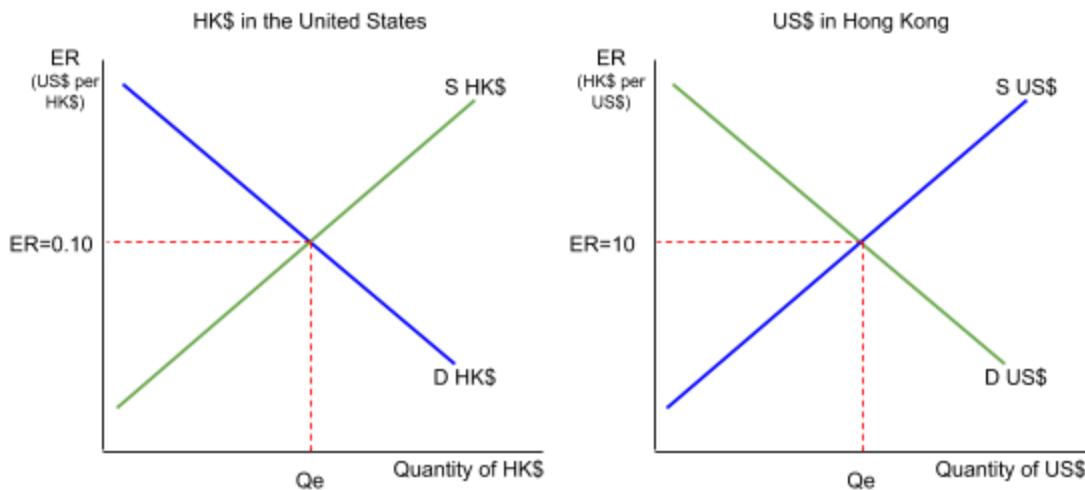
A **fixed exchange rate system** exists when a currency's value against one or more other currencies is set by the government or central bank in order to promote particular macroeconomic objectives. Exchange rate fixing requires governments or central banks to intervene in the forex market to manipulate the demand for or supply of the currency.

A **managed exchange rate system** is when a currency's value against one or more other currencies is allowed to fluctuate between a certain range by the country's government or central bank. If the exchange rate gets below a certain level or above a certain level, then the government or central bank will intervene to bring it back within the desired range.

Fixed exchange rate systems

A country may chose to fix, or **peg**, its currency against that of a major trading partner. For instance, assume the government of Hong Kong wishes to avoid volatility in the value of the HK dollar, thus it decides to peg the HK dollar against the US dollar. Assume the free market rate is currently \$10 HK to \$1 US (or \$0.10 US per \$1 HK). Hong Kong wishes to reduce the value to \$12 HK per \$1 US (or \$0.083 US per \$1 HK).

The graphs below show the value of the HK\$ in the US and the US\$ in Hong Kong before any intervention in the forex markets.



Hong Kong wishes to **devalue** the HK\$ against the US\$. This means it will take action in the forex market to weaken the currency. To do so, it has three options:

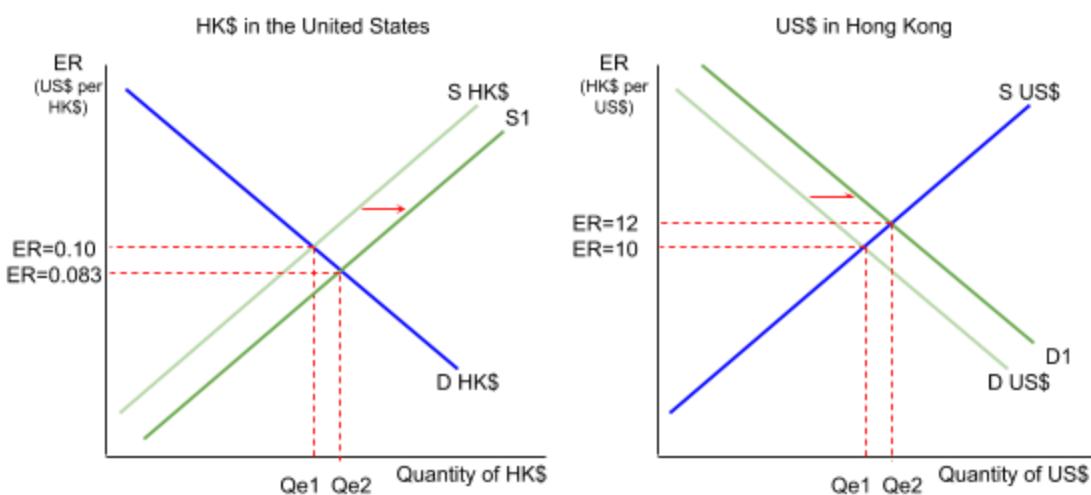
- **Lower interest rates:** Through expansionary monetary policy, Hong Kong's central bank can reduce demand for HK\$ by reducing the interest rate in Hong Kong. This will lead to an outflow of foreign investment in Hong Kong as assets there become

less attractive due to the lower interest rates. Demand for HK\$ will decrease, supply of US\$ will decrease, and the HK\$ will lose value while the US\$ gains value.

- **Forex market intervention:** Hong Kong's central bank could directly intervene in the forex markets and increase the supply of HK\$ by buying up US\$. This would cause the HK\$ to lose value and the US\$ to gain value, achieving the government's desired outcome of a devalued HK\$.
- **Exchange controls:** Hong Kong's government could place limitations on the amount of foreign investment in Hong Kong, essentially limiting the demand for HK\$ and limiting the supply of US\$. Exchange controls would cause the HK\$ to lose value and the US\$ to gain value.

The three options above would all result in a **devaluation** of the HK\$ against the US\$. Once the desired exchange rate is achieved (\$12 HK per US\$ and \$0.083 US per HK\$), the government and central bank of Hong Kong must continually intervene using the three tools above to maintain the peg.

The graph below show the effect of the forex market interventions on the values of the HK\$ and the US\$.



Hong Kong's intervention in the forex market has **devalued** the HK\$ to a rate lower than its free market equilibrium, which may benefit Hong Kong's economy through making exports cheaper and boosting aggregate demand.

When a country wishes to make its currency stronger on the forex market, it will engage in a **revaluation**. Methods for revaluing a currency include:

- **Increasing the interest rate:** Through contractionary monetary policy the central bank can increase demand for HK\$ by making investments in the territory more attractive to foreign investors. Higher interest rates will increase demand for HK\$ and increase supply of US\$, causing the HK\$ to increase in value.
- **Forex market intervention:** Hong Kong's central bank could directly intervene in

the forex market by purchasing HK\$ using its reserves of US\$. The increased demand for HK\$ would cause it to gain value while the increased supply of US\$ would cause it to weaken.

- **Loosening exchange controls:** Assuming Hong Kong has exchange controls in place, loosening them would let more foreign investment flow into Hong Kong and cause the HK\$ to get stronger on the forex market.

Any of the above interventions would cause demand for HK\$ to increase and supply of US\$ to increase, causing resulting in a **revalued** HK\$. A stronger HK\$ would make imports cheaper and could be used to fight inflation, essentially acting as a contractionary demand-side policy.

Managed exchange rate systems

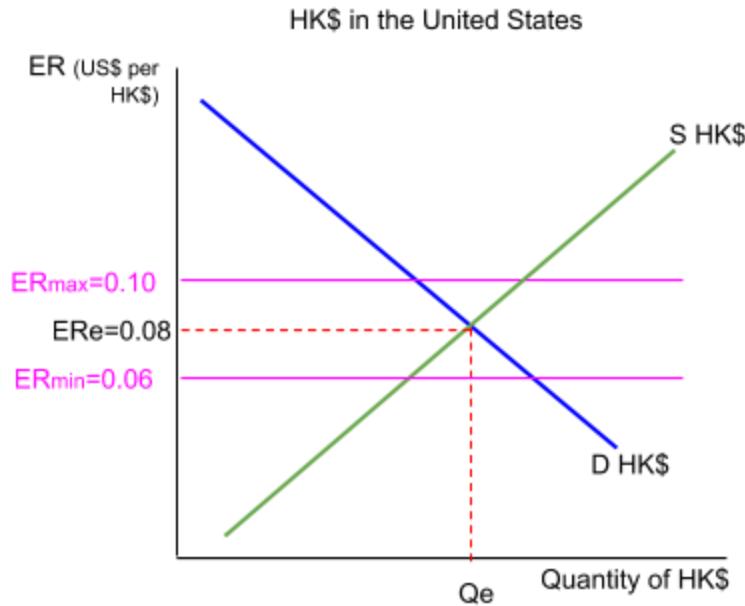
- Explain how a managed exchange rate operates, with reference to the fact that there is a periodic government intervention to influence the value of an exchange rate.
- Examine the possible consequences of overvalued and undervalued currencies.

Fixed exchange rate regimes are rather rare, largely due to the perception they create among trading partners that they act as a form of protectionism, sheltering domestic producers from foreign competition by undervaluing the currency. More common are **managed exchange rate systems**, also called **managed floats**, through which the government and central bank set a target range of exchange rates between which the currency's value is allowed to float.

If a currency's market value is pushed above the “ceiling rate,” the central bank or government will intervene to devalue the currency. If its market value falls below the “floor rate,” intervention will be undertaken to revalue the currency to keep within the desired range.

Managed floats result in a more stable and predictable exchange rate, promoting macroeconomic stability and a favorable environment for foreign investment, possibly promoting foreign demand for the country's exports.

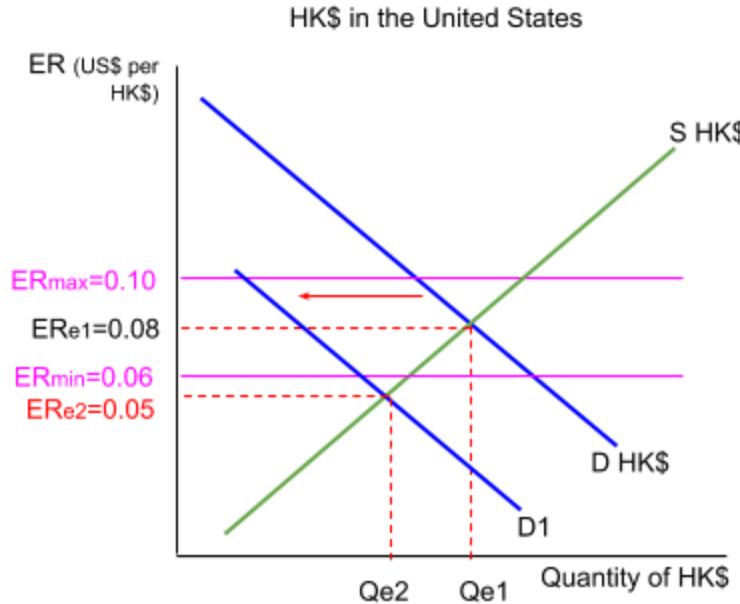
Graphically, a managed float involves the use of price controls, as seen in the graph below.



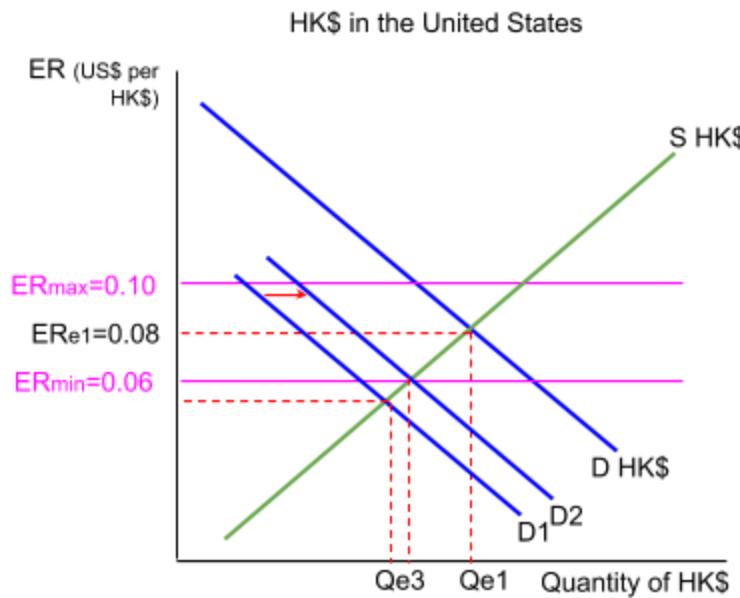
Observe from the graph:

- The current equilibrium exchange rate of \$0.08 US per HK\$ is within the desired range of \$0.06 to \$0.10.
- If an increase in demand for HK\$ or a decrease in supply of HK\$ causes the market rate to rise above \$0.10 US, the central bank will intervene by lowering interest rates or increasing the supply of HK\$ to devalue it and bring it back below \$0.10 US per HK\$.
- If a decrease in demand for HK\$ or an increase in supply of HK\$ causes the market rate to fall below \$0.06 US per HK\$, the central bank will intervene by raising interest rates or buying up HK\$ to revalue it and bring it back above \$0.06 US per HK\$.

For example, assume that due to a decrease in foreign demand for shares in companies listed on the Hong Kong stock exchange, the demand from US investors for HK\$ decreases, driving the equilibrium exchange rate down to \$0.05 US per HK\$, as in the graph below.



In order to maintain the “floor rate” of \$0.06 US per HK\$, Hong Kong’s central bank must raise interest rates to boost demand for investment from the US. If that does not work, direct intervention in the forex market through a purchase of HK\$ by Hong Kong’s central bank will be necessary. Either of these two interventions will increase demand for HK\$ and bring the exchange rate back within the target range.



Demand has increased from D₁ to D₂ due to Hong Kong’s central bank actions. The HK\$ is back within its target range of \$0.06 US to \$0.10 US.

Evaluation of different exchange rate systems

- Compare and contrast a fixed exchange rate system with a floating exchange rate system, with reference to factors including the degree of certainty for stakeholders, ease of adjustment, the role of international reserves in the form of foreign currencies and flexibility offered to policy makers.

Different exchange rate systems have their advantages and disadvantages. Essentially it all comes down to how tolerant policymakers are to the uncertainty that may accompany a floating exchange rate.

Some of the advantages of a managed exchange rate (and thus the disadvantages of a floating exchange rate) are:

- Greater stability and **certainty for international investors** considering investing in the country's financial and real assets.
- More **stability for the country's export firms** that depend on demand from overseas markets.
- More **certainty for the nation's consumers** and firms who may depend on imported products.
- The **ability to promote macroeconomic objectives** by stimulating or contracting aggregate demand through devaluing or revaluing the currency.
- A **competitive trade advantage** can be enjoyed by a country that keeps its currency artificially weak, or undervalued against other currencies.
- A country with a managed exchange rate will have a **reduced risk of speculation** on its forex market. Under a floating exchange rate, if investors expect the country's currency to appreciate, it will likely do so and harm the nation's producers and reduce aggregate output. Management prevents such speculative shocks to the exchange rate.

Some disadvantages of managed exchange rate systems (and advantages of a floating exchange rate) include:

- Monetary policy must constantly be adjusted to maintain the exchange rate, and therefore is not primarily used to promote price level stability or other domestic macroeconomic objectives. A floating exchange rate allows for **monetary policy freedom** and the ability to focus on domestic macro objectives.
- Intervening in the forex markets requires central banks to maintain reserves of foreign exchange that otherwise could be used by domestic consumers and producers to purchase imports. With a floating exchange rate system, central banks can keep **less foreign reserves** for use in exchange rate management.
- Undervalued currencies tend to result in imbalanced trade and **persistent trade surpluses** with other countries, which could result in retaliation by trading partners upset with a country's devaluation of its currency.
- Overvalued currencies tend to result in **persistent trade deficits** and disadvantage the country's producers. In contrast.

Floating exchange rates respond to changes in supply and demand and tend to automatically

adjust to trade flows, resulting in **relatively balanced trade** between a country and its trading partners (assuming they also allow their currencies to float).

3.3 The balance of payments

The meaning of the balance of payments

- Outline the role of the balance of payments.
- Distinguish between debit items and credit items in the balance of payments.

Introduction to the balance of payments (BoP)

Through international trade, nations are constantly exchanging goods, services and financial and real assets across national borders. A country's **balance of payments (BoP)** measures the flow of money for financial and real transactions between the households, businesses, banks, and the government of one nation and all other nations.

The BoP includes two "accounts", which are basically just tables recording the money flowing into and out of a country for different transactions with the rest of the world. The current account measures the flow of goods and services and income transfers, while the capital and financial account measures the transfers of capital and the flow of funds for investment in financial and real assets between countries.

The current account (CA)

A country's **current account** measures the balance of trade in goods and services and the flow of income between one country and all other countries. It also records monetary gifts or grants that flow into or out of a country.

The current account can be broken down into several components:

- The **goods balance** (also called the visible balance) measures the spending by consumers and firms in one country on other nations' goods (both consumer and capital goods) as well as spending by consumers in the rest of world on the recording nation's goods. A country's visible trade balance includes credits and debits.
 - **Credits (+):** Money earned from exports is recorded as a positive in the current account and moves the CA balance towards surplus
 - **Debits (-):** Money spent on imports is recorded as a negative in the current account and move the CA balance towards deficit
- The **services balance** (also called the invisible balance) measures spending on non-tangible purchases such as tourism, banking, consulting, legal services, and transportation. Services can be "imported" and "exported", although there will be no physical transportation of a product involved.
 - **Credits (+):** Spending on services by foreigners is recorded as a positive and move the CA balance towards surplus.
 - **Debits (-):** Spending on services provided by foreigners and consumed by domestic households is recorded as a negative and move the CA balance towards deficit.

In addition to trade in goods and services, the current account also measures the flow of

income and transfers.

- The **income balance** records the transfer of incomes earned by citizens of one country from activities in another country back to the income earner's country of origin are also measured in the current account. This includes the wage income earned by a country's citizens for employment by foreign companies abroad.
 - **Credits (+):** Income earned abroad and sent home is recorded as a positive and moves the current account towards surplus
 - **Debits (-):** Income earned at home by foreigners and sent abroad is recorded as a negative and moves the current account towards deficit.
- The **transfers balance** records payments made from one nation to another that are not in exchange for any good or service, such as gifts or grants. Transfers are divided into two categories, official transfers are payments from one government to another, sometimes known as "aid", and private transfers are payments made by citizens of one country to residents of any other country.
 - **Credits (+):** Money transferred from a foreign nation into the home country is recorded as a positive and moves the current account towards surplus.
 - **Debits (-):** Money transferred from the home country to foreign countries is recorded as a negative and moves the current account towards deficit.

The capital and financial account (CFA)

The other major account measured in the BoP is the **capital and financial account**, which measures the exchanges between a nation and the rest of the world involving ownership of financial and real assets, including:

- **Investment in physical assets:** Foreigners may buy and sell a country's physical assets, including real estate, factories, office buildings and other factors of production
 - **Credits (+):** When foreigners invest in a country's physical assets (for example, a foreign automaker builds a plant in Ohio), it is recorded as a positive and the CFA moves towards surplus.
 - **Debits (-):** When domestic firms invest in physical assets in a foreign country (for example, when an American tech company builds a data center in India), it is recorded as a negative for the investing country and the CFA moves towards deficit.
- **Financial investment:** International purchases of financial assets, such as shares in companies and government or corporate debt (bonds) bonds are also measured in the financial account.
 - **Credits (+):** When foreign investors invest in the domestic country's stock or bonds markets, it is recorded as a positive and moves the CFA towards surplus.
 - **Debits (-):** When domestic investors invest in foreign stock or bond markets, it is recorded as a negative and moves the CFA towards deficit.
- **Capital transfers:** A sub-account of the financial account is the capital account, which measures the transfer of capital goods, money for the purchase of capital goods, and debt forgiveness between one nation and others.
 - **Credits (+):** Money transferred by the citizens of foreign countries for

- investments in physical or real assets, for debt forgiveness or for capital acquisition move the financial account balance towards a surplus
- **Debits (-):** Money transferred from the home country abroad for investments in physical or real assets, for capital acquisition or for debt forgiveness in a foreign country move the financial account balance towards a deficit.

The components of the balance of payments accounts

- Explain the four components of the current account, specifically the balance of trade in goods, the balance of trade in services, income and current transfers.
- Distinguish between a current account deficit and a current account surplus.
- Explain the two components of the capital account, specifically capital transfers and transaction in non-produced, non-financial assets.
- Explain the three main components of the financial account, specifically, direct investment, portfolio investment and reserve assets.

HL only objective:

- Calculate elements of the balance of payments from a set of data.

The sum of the four sub-accounts (goods, services, income and transfers) gives us the **current account balance**. A country's current account can be in either:

- **Deficit:** When a nation's current account balance is negative. The country spends more on imported goods and services, income and transfers to the rest of the world than it receives in payments for goods and service exports, income and transfers.
- **Surplus:** When a nation's current account balance is positive. The country earns more from its sale of exported goods and services, income and transfers from the rest of the world than it makes in payments for other countries' exports and income and transfers.

The table below shows the balance on each component of New Zealand's current account, along with the final current account balance, for 2010.

Account	Credits (millions of NZ\$)	Debits (millions of NZ\$)	Balance (millions of NZ\$)
Goods	29,109	-29,706	-597
Services	11,966	-9,777	2,189
Income	2,844	-8,851	-6,007
Current Transfers	1,318	-1,128	190
		Current Account Balance	-4,225

The capital and financial account balance

The sum of the three categories of investment and capital transfers determines a country's CFA balance, which can be either in:

- **Surplus:** If the flow of money into a nation's financial account is greater than the flow out of the country in a particular year, the country's financial account is in surplus.
- **Deficit:** If the flow of money out of a country for financial transaction is greater than the flow into the country in a particular year, the country's financial account is in deficit

Official foreign exchange reserves

If in a given year, the balance of a nation's current and financial account is not equal to zero, then the difference will be added or subtracted from the nation's official reserves of foreign exchange. Consider New Zealand's current and financial account balances:

Account	Credits (millions of NZ\$)	Debits (millions of NZ\$)	Balance (millions of NZ\$)
Goods	29,109	-29,706	-597
Services	11,966	-9,777	2,189
Income	2,844	-8,851	-6,007
Current Transfers	1,318	-1,128	190
		Current Account Balance	-4,225

Account	Credits (millions of NZ\$)	Debits (millions of NZ\$)	Balance (millions of NZ\$)
Direct Investment	3,895	-1,293	2,602
Portfolio Investment	3,920	-6,947	-3,027
Other Investments	1,272	289	1,561
Capital transfers	1,576	-814	762
		Financial Account Balance	1,898

The CA deficit is larger than the CFA surplus by \$2,327. This is the amount by which the

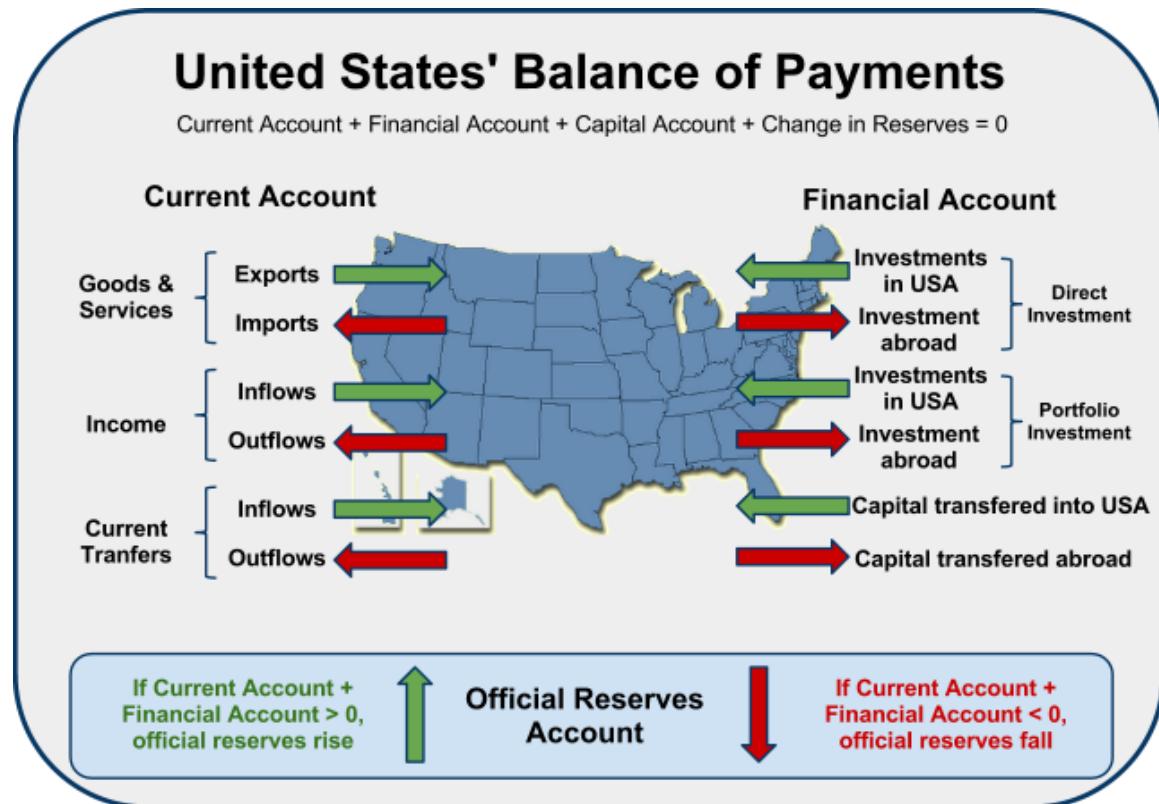
outflow of funds in the current account exceeds the inflow of funds in the financial account. This amount ends up in foreign central banks and is considered an asset of New Zealand held by foreigners.

- A net deficit in the CA and CFA (as in the case of New Zealand) actually results in an inflow (thus, a positive sign) in the official reserves account, since the deficit country must sell its reserves of foreign currency to make up for the net deficit.
- If a country has a net balance of payment surplus, then the change in the foreign exchange reserves is recorded as a negative since the country's ownership of assets denominated in foreign currencies actually increases each year its net CA and CFA account balance is positive.

The relationship between the accounts

- Explain that the current account balance is equal to the sum of the capital account and financial account balances.
- Examine how the current account and the financial account are interdependent.

The graphic below provides a summary of the inflows and outflows in a nation's balance of payments.



Observe from diagram:

- The green arrows pointing inwards indicate money flowing into the country.
- The red arrows pointing outwards indicate money flowing out of the country. The total balance of payments must always be equal to ZERO.

- A current account deficit will be offset by a financial account surplus
- A current account surplus will be offset by a financial account deficit

The sum of all the accounts in the Balance of Payments will equal ZERO

$$\text{Current account} + \text{Financial Account} + \text{change in the official reserves} = 0$$

The relationship between the current account and the exchange rate

- Explain why a deficit in the current account of the balance of payments may result in downward pressure on the exchange rate of the currency
- Explain why a surplus in the current account of the balance of payments may result in upward pressure on the exchange rate of the currency.

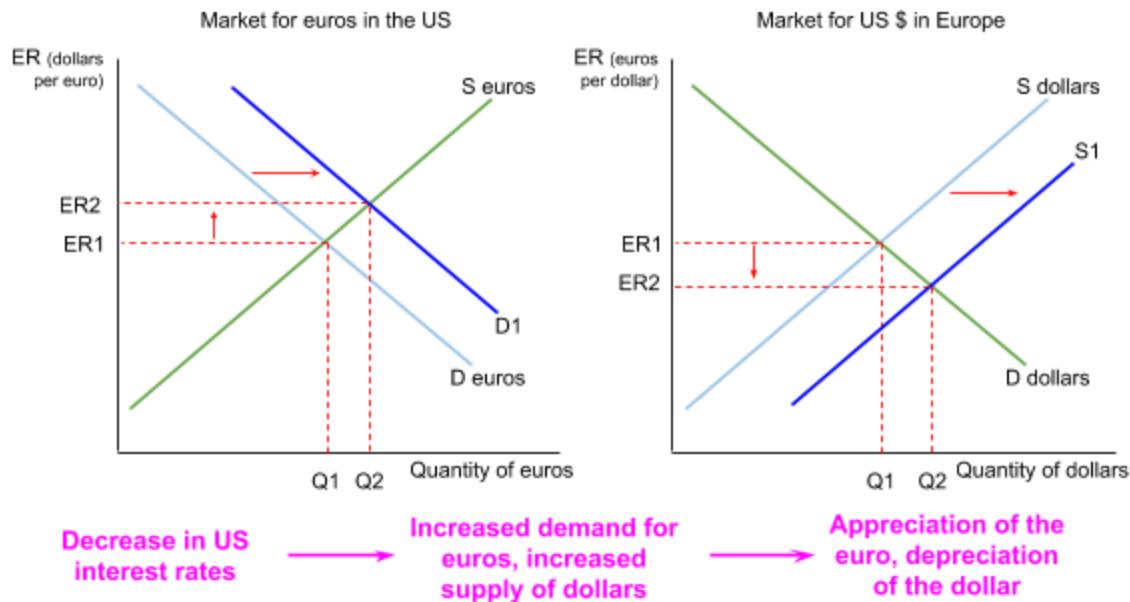
Under a **floating exchange rate system**, countries' currencies are allowed to appreciate and depreciate free of government or central bank manipulation as demand and supply on forex markets change due to international trade and investment flows.

Changes in the exchange rate, in the long run, will resolve imbalances in the current and capital/financial accounts. In other words, if a country's currency is allowed to float freely, countries will not have persistent trade deficits or surpluses.

Changes in the exchange rate lead to changes in the BoP

For example, let's assume that the US and Europe, which are major trading partners, currently experiencing balanced trade. The two economies have current account balances with one another of zero (e.g. exports equal imports). How will a change in the exchange rate disrupt the flow of goods and services between the US and Europe?

Assume the US central bank (the Federal Reserve) lowers US interest rates. The effects of an expansionary monetary policy in the US on the forex markets for the dollar in Europe and the euro in the US are shown below.



Following the decrease in US interest rates:

1. Demand for euros in the US increases as European investments now have a relatively higher interest rate. **The euro appreciates.**
2. Supply of dollars in Europe increases as more Americans convert their dollars to euros to invest in European assets. **The dollar depreciates.**
3. **Europe's CFA moves towards surplus and America's CFA moves towards deficit** as funds for investment flow from the US to Europe.
4. The stronger euro leads Europeans to demand more American exports and Americans to demand fewer European imports. **America's CA moves towards surplus and Europe's CA moves towards deficit.**

A change in a factor affecting the dollar/euro exchange rate (a fall in US interest rates) has impacted both the capital/financial accounts and the current accounts in the US and Europe. The weaker dollar has caused America's net exports to increase and Europe's to increase. America's CA has moved towards surplus and its CFA towards deficit.

Over time, European's greater demand for American goods will cause demand for dollars to rise and the dollar to appreciate against the euro. As the dollar appreciates, European consumers will begin demanding fewer US goods, and Americans will demand more European goods.

In the long run, the floating exchange rate should lead to relatively balanced trade between the US and Europe. The result of floating exchange rates is balanced trade over time between two economies!

As we demonstrated in the last section a change in interest rates in one country will affect capital flows between that country and the rest of the world, assuming all else is held constant.

Recall that the **interest rate** is the opportunity cost of spending money; or we could think of it as the benefit of saving money. When a country's real interest rate increases, *ceteris paribus*, other countries investors will wish to save more money in that country.

- An **increase in interest rates**, therefore, causes an inflow of foreign capital to a country, causing the country's currency to appreciate.
- A **decrease in interest rates** causes an outflow of capital into foreign economies as domestic investors wish to put their money in foreign assets, causing the lower interest rate country's currency to depreciate.

Recall also that a change in a country's capital/financial account balance is always accompanied by an opposite change in its current account balance. The table below shows the effect of interest rate changes on a country's international capital flows, the exchange rate, and the balance of payments.

Change in interest rates	Capital flows	Exchange rate	CFA balance moves towards...	CA balance moves towards...
Increase	inflow	appreciate	surplus	deficit
Decrease	outflow	depreciate	deficit	surplus

Central bank policy and exchange rates

Because changes in the interest rate can affect exchange rates, international capital flows, and the balance of payments, central banks can wield their power over the interest rate to achieve international trade objectives, in addition to the typical macroeconomic objectives outlined earlier in the course.

For example, assume a country has experienced persistent current account deficits with several of its trading partners, resulting in weak aggregate demand and increasing structural unemployment. The central bank could choose to intervene in order to both stimulate domestic demand, but also to increase net exports and thus reduce the current account deficit. An **expansionary monetary policy** will help resolve some of this country's economic problems.

- Lower interest rates cause a decrease in foreign investment and a **capital outflow**.
- Less demand for the currency causes it to **depreciate** on forex markets.
- A weaker currency makes the country's exports more attractive to foreigners and shifts domestic consumers away from now more expensive imports and towards domestically produced goods. **Net exports increase and the current account deficit shrinks.**

Implications of persistent current account deficit (HL only)

- Discuss the implications of a persistent current account deficit, referring to factors including foreign ownership of domestic assets, exchange rates, interest rates,

indebtedness, international credit ratings and demand management.

A deficit in the balance of trade between a nation and all others can have some negative short-run and long-run consequences on the nation's economy. Most notably, a CA deficit means a lower level of aggregate demand for the nation.

Assume a country producing at full employment experiences an increase in its current account deficit:

- Domestic consumption of foreign goods increases and foreign consumption of domestic goods decreases, causing net exports, a component of AD, to fall.
- The fall in AD causes a decrease in the country's output, employment, and price level.
- The economy experiences a recession.

Besides the impact on domestic employment and aggregate demand, the following consequences may result for a nation experiencing a persistent deficit in its current account.

Consequences of a persistent current account deficit	
Depreciation of the currency	A primary determinant of exchange rates is the demand for exports and imports. If domestic consumers demand more imports than foreigners demand of the home country's exports, then the value of the domestic currency will fall relative to foreign currencies. A weaker currency makes imported raw materials more expensive and can contribute to cost push inflation.
Increased foreign ownership of domestic assets	Since current and capital accounts must be in balance, a deficit in the current account means a country likely has a surplus in its capital account. This means foreigners own more of the home country's assets (factories, land, government debt, company shares, etc...) than domestic investors own of foreign assets. Such foreign ownership of domestic assets may pose a threat to the economic sovereignty (freedom) of the deficit country.
Higher interest rates	In order to offset the inflationary effects of a weak currency, a country's central bank may try to strengthen the currency by raising interest rates to attract foreign capital to the country. A higher interest rate will negatively affect domestic investment by firms, slowing growth in the nation's capital stock over time.
Increased indebtedness	A current account deficit is offset by a financial account surplus. One of the domestic assets foreign investors will demand is government bonds. Increased selling of bonds by the government to foreign investors increases the amount of national debt held by foreigners. When a government has large amounts of foreign held debt, it must pay interest on that debt, meaning taxpayer money is being paid to foreigners, reducing the government's ability to spend as much on domestic projects like infrastructure, education and health care.

Methods to correct a persistent current account deficit (HL only)

- Explain the methods that a government can use to correct a persistent current account deficit, including expenditure switching policies, expenditure reducing policies and supply-side policies, to increase competitiveness.

- Evaluate the effectiveness of the policies to correct a persistent current account deficit

The government or central bank of a nation experiencing a persistent current account deficit may choose to intervene to avoid the negative consequences of the deficit on the nation's economy.

Methods for correcting a current account deficit	
Exchange rate devaluation	By intervening in the foreign exchange market a government or central bank can try and devalue the country's currency. Trading its own currency for foreign currencies will increase the currency's supply, causing it to depreciate. Greater demand for other countries' currencies will cause them to appreciate. These changes should increase the country's net exports.
Expansionary monetary policy	Expansionary monetary policies would lower the country's interest rates and lead to an outflow of capital in the financial account, depreciating the currency and reducing a trade deficit
The use of protectionism	New tariffs or quotas will raise the cost of imported goods and allow domestic producers to sell their products at a higher price at home. Subsidies for domestic producers reduce domestic costs of production and therefore reduce the price of domestic goods, making imports less attractive
Contractionary demand-side policy	An increase in taxes or a decrease in government spending will reduce aggregate demand and the average price level, reducing national income and spending on imports. The lower price level will make exports more attractive to foreigners, bringing the current account into balance.
Supply-side policies	Market-based supply side policies can reduce a nation's costs of production and make its producers more competitive with foreign competitors. Lower costs and prices would lead domestic consumers to demand more domestically produced goods and fewer imports, and at the same time increase foreign demand for exports, reducing a current account deficit in the process.

The Marshall-Lerner condition and the J-curve effect (HL only)

- State the Marshall-Lerner condition.
- Apply the Marshall-Lerner condition to the effect of devaluation/depreciation on the current account.
- Explain the J-curve effect, with reference to the Marshall-Lerner condition.

Whether or not a depreciation of a nation's currency will cause its current account to move towards deficit depends on the price elasticity of demand for exports and imports.

The **Marshall-Lerner Condition (MLC)** says that if the combined PEDs for exports and imports are greater than one (e.g. elastic) then a depreciation of the currency will cause the current account balance to move towards surplus.

- If the MLC is met, then a country can successfully reduce a current account deficit by devaluing its currency.
- If the MLC is NOT met, then currency devaluation will actually worsen a current

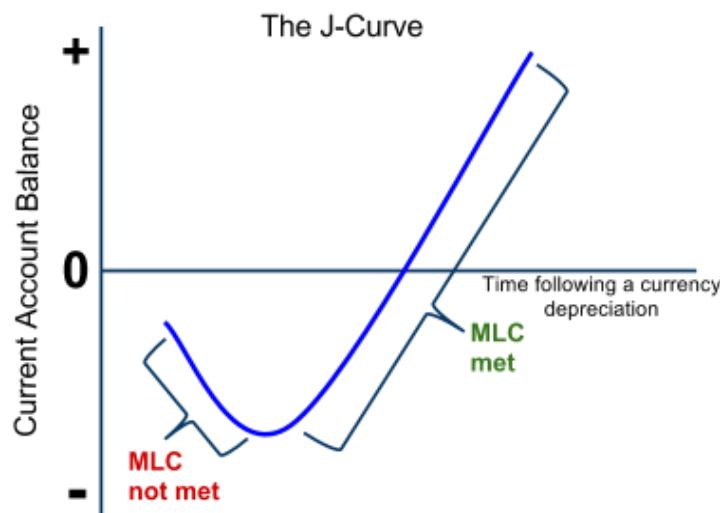
account deficit.

Consumers must be relatively responsive to the changing prices of imports and exports in order for a currency devaluation to be effective. For instance,

- If demand for a nation's exports is highly inelastic, then their cheaper prices abroad (following a devaluation of the currency), will actually result in foreigners spending less on the nation's goods (since they will have to give up less of their own currency to buy them).
- If demand for imports is highly inelastic, then a devaluation result in domestic consumers actually spending more on imports, since they are now more expensive and the quantities demanded have not decreased by much.
- In this situation, the MLC is NOT met and a weaker currency will lead to a larger trade deficit.

The J Curve

Recall from microeconomics that consumers tend to be more responsive to price changes as time passes following the change in price. Based on this knowledge, we can predict what will happen to a nation's current account balance over time following a depreciation of its currency.



Assume the country was experiencing a current account deficit:

- The central bank decides to devalue the currency by lowering interest rates and increasing its supply on forex markets.
- In the first several weeks following the devaluation, consumers at home and abroad are relatively unresponsive to the country's now higher priced imports and lower priced exports. The MLC is not met, so the current account deficit actually worsens.
- However, after a couple of months, consumers at home and abroad have begun to respond to the country's cheaper exports and the more expensive imports from abroad. The MLC is now met and the current account begins to move towards surplus.

Currency devaluations will be ineffective at correcting current account deficits in the short-run. But over time, a weaker currency will likely move a nation's current account towards surplus as consumers become more responsive to the higher prices of imports and the lower prices of exports!

Implications of a persistent current account surplus (HL only)

- Discuss the possible consequences of a rising current account surplus, including lower domestic consumption and investment, as well as the appreciation of the domestic currency and reduced export competitiveness.

A current account surplus occurs when a nation's exports are greater than its imports over a period of time. A persistent current account surplus has some positive effects for a nation, but also some negative ones.

Consequences of persistent current account surpluses	
Appreciation of the currency	Since a current account surplus means the country is exporting more than it is importing, foreigners are demanding more of the surplus nation's currency, putting upward pressure on the value of the exchange rate. An appreciating currency will harm producers in the export sector and could reduce domestic employment
Increased ownership of foreign assets	A surplus in the current account is usually offset by a deficit in the financial account. Domestic investors will increase their ownership of foreign assets (stocks, government debt, real estate and factories), meaning there is a net outflow of capital from the country
Reduced levels of domestic consumption	If a nation exports a large proportion of its total output, there is less stuff left over for domestic households to consume. While a current account surplus may be good for employment, it is often bad for domestic consumption, since the money earned from exported goods and services is not entirely spent on imports
Possibility of increased protectionism	Foreign governments unhappy with the trade imbalance with the surplus nation may threaten to impose protectionist measures on the exporting nation's goods. Such measures will undermine the surplus nation's comparative advantage and reduce employment and output

3.4 Economic Integration

Introduction to economic integration

When two or more countries join formal agreements through which they commit to reducing or eliminating protectionist trade barriers on a group of products or all products, they are participating in **economic integration**. Economic integration takes many forms, from very low level agreements through which countries reduce tariffs on a small number of goods to complete economic integration, through which countries agree to completely free trade, free movement of labor and capital, and even share a common currency and monetary and fiscal policies.

Economic integration is a controversial issue in the 21st century, and in 2018, for the first time since the end of World War 2, more trade barriers are being erected between the world's largest trading nations than are being torn down. Under United States president Donald J. Trump tariffs have been placed on every one of America's largest trading partners and the United States has either withdrawn from or threatened to withdraw from some of its most significant free trade agreements. In response, America's trading partners have retaliated with new tariffs of their own on US goods.



The **trade wars** of 2018 represent a marked about turn on the path towards increased economic integration of the last several decades. In the coming sections, we'll examine the different forms of economic integration, outline their advantages and disadvantages, and look at who wins and who loses from freer trade between the world's economies.

Preferential trade agreements

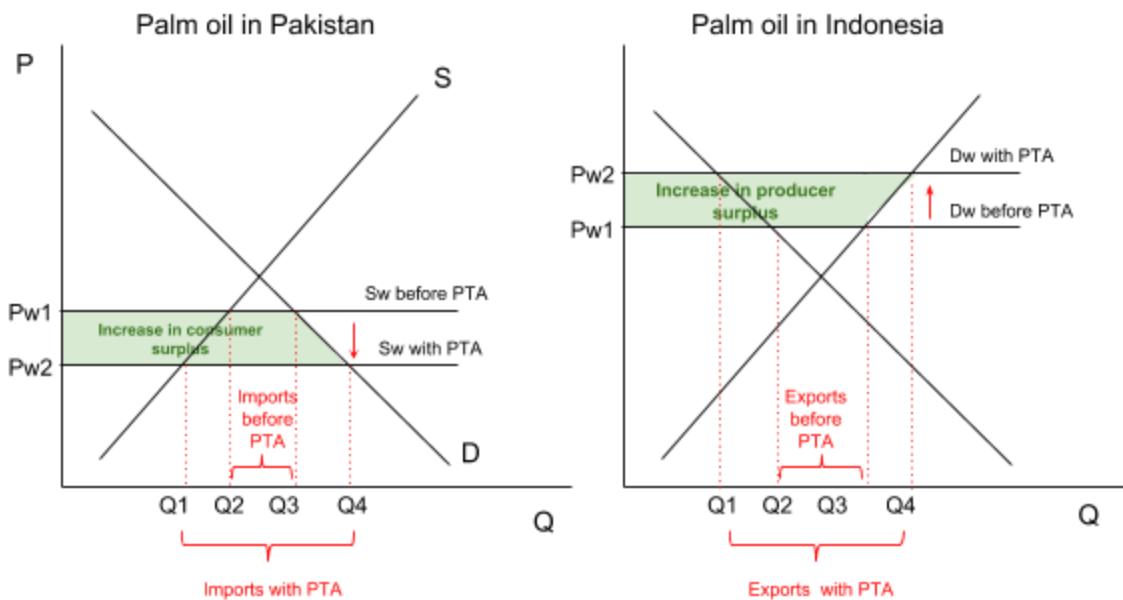
- Distinguish between bilateral and multilateral (WTO) trade agreements.
- Explain that preferential trade agreements give preferential access to certain products from certain countries by reducing or eliminating tariffs, or by other agreements relating to trade.

The lowest form of economic integration is a **preferential trade agreement (PTA)**, which is when two countries (a bilateral agreement) or more than two countries (a multilateral agreement) agree to reduce or eliminate tariffs on particular products. PTAs may also require elimination of quotas and subsidies on the products to which they relate.

The World Trade Organization (WTO) is an example of large, multi-lateral preferential trade agreement through which the 159 member countries (as of 2018) agree to the reduction of tariffs on most goods traded among member countries.

An example of a bilateral PTA is the agreement between Pakistan and Indonesia that includes several agricultural products, including palm oil. To increase the overall production

and lower the prices of palm oil, Pakistan agreed to eliminate all tariffs on Indonesian palm oil. The effect of the PTA in the markets for palm oil in Pakistan and Indonesia can be seen below.



Observe from the graphs:

- The removal of tariffs on Indonesian palm oil causes the price of oil in Pakistan to fall from P_w2 to P_w1 .
- Lower prices lead to more oil consumption in Pakistan ($Q3$ to $Q4$) and an increase in consumer surplus represented by the green area.
- Increased demand from Pakistan causes the world demand as seen by Indonesia increase from "Dw before PTA" to "Dw with PTA", increasing the price in Indonesia from P_w1 to P_w2 .
- A higher price leads Indonesian firms to increase production from $Q3$ to $Q4$, leading to an increase in producer surplus represented by the green area.
- Pakistan imports more palm oil, Indonesia exports more palm oil, and there is an increase in total surplus in both countries.

The PTA has increased overall welfare in both Pakistan and Indonesia and resulted in a higher total level of palm oil production and a more efficient allocation of resources. Pakistan will use less of its scarce land to produce a product that it is not particularly good at producing, while Indonesia will allocate more land to palm oil production, in which it has a comparative advantage.

Of course, some stakeholders are worse off because of the PTA.

- Pakistani producers sell less oil at a lower price as imported oil crowds out domestic production.
- Indonesian consumers will buy less domestic oil at a higher price, as exports crowd out domestic consumption.

Trading blocs

- Distinguish between a free trade area, a customs union and a common market.
- Explain that economic integration will increase competition among producers within the trading bloc.
- Compare and contrast the different types of trading blocs

HL only objectives:

- Explain the concepts of trade creation and trade diversion in a customs union.
- Explain that different forms of economic integration allow member countries to gain from economies of scale.

PTAs are just the lowest level of economic integration. When two or more countries agree to eliminate tariffs on ALL goods traded between them, they have joined what is known as a **trading bloc**. There are three kinds of trading blocs, ranked below from the lowest to the highest level of economic integration.

Trading blocs		
Type	Includes..	Examples
Free trade area	removal of tariffs on all goods (and some services)	The North American Free Trade Agreement (NAFTA - Canada, the United States, and Mexico)
Customs union	removal of tariffs on all goods (and some services), and... common external tariffs on goods from non-member countries	Southern African Customs Union (SACU - Southern Africa: Botswana, Lesotho, Namibia, South Africa, and Swaziland)
Common market	removal of tariffs on all goods and services, and... common external tariffs on goods from non-member countries, and... free flow of labor and capital between member countries	The East African Community (EAC - Burundi, Kenya, Rwanda, South Sudan, Tanzania, and Uganda)

Notice that each level of trade bloc builds on the lower level. A customs union IS a free trade area in which there are common external tariffs. A common market IS a customs union in which there is free flow of capital and labor.

Effects of trade blocs

Economic integration through the removal of tariffs and increased movement of labor and capital will have several effects on various stakeholders in the member countries:

- Increased competition between producers
- Lower prices for consumers
- Greater efficiency

Trade creation versus trade diversion (HL only)

When countries join a customs union, many of the benefits of freer trade may be enjoyed between the member countries, but it is also possible that a less efficient allocation of resources could result from the implementation of common external tariffs on non-member countries.

For example, consider the Southern African Customs Union, which includes Botswana, Lesotho, Namibia, South Africa, and Swaziland (see map below).



Let's assume that before the formation of the customs union, the five member countries all had their own tariffs on one another's goods and on other countries' goods, such as Zimbabwe (which is not part of the union). Assume also that Zimbabwe is the most efficient producer of cotton in southern Africa. The other countries all produce cotton as well, but at a higher cost per unit than Zimbabwe.

With the formation of the customs union, the five member countries agree to remove tariffs on cotton between one another, but agree on and establish a common tariff on cotton from Zimbabwe. Assume that of the five member countries, South Africa produces cotton most

cheaply. As a result of the customs union, South Africa will produce more cotton and export it to the other four, **creating trade** within the customs union. However, less cotton will be imported from Zimbabwe, as trade is **diverted** from a low cost country (Zimbabwe) to a higher cost country (South Africa).

Trade creation refers to the effect of a customs union that results in a good's production shifting from a high cost country to a low cost country. When production shifts from Namibia, Botswana, Swaziland, and Lesotho to South Africa, trade is created.

Trade diversion refers to the effect of a customs union that results in a good's production shifting from a low cost country to a high cost country. When production shifts from Zimbabwe to south Africa as a result of the common external tariffs on Zimbabwe's cotton, trade is diverted.

Trade creation increases allocative efficiency within member countries of a customs union. Trade diversion reduces allocative efficiency, as production shifts from more efficient producers to less efficient producers.

Monetary union

- Explain that a monetary union is a common market with a common currency and a common central bank.
- Discuss the possible advantages and disadvantages of a monetary union for its members.

The highest level of economic integration in which sovereign countries can participate is **monetary union**, which is created when members of a common market share a single currency, a single central bank, and monetary policy.

The most prominent monetary union is the Eurozone, which includes 19 of the 28 countries in the European Union (EU), which in turn are part of the European Economic Area (EEA), a common market which includes the 28 EU countries plus Iceland, Liechtenstein, and Norway.

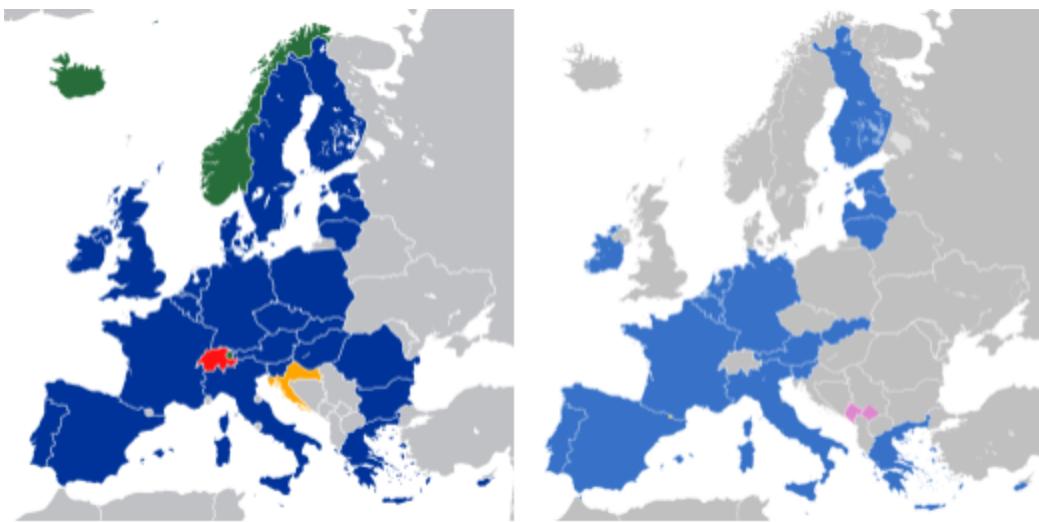


Figure 1: European Economic Area common market (blue and green countries)

Figure 2: Eurozone monetary union (blue countries)

Eurozone countries share a single currency, the euro, the supply of which is controlled by a single central bank, the European Central Bank.

Advantages of a monetary union

Joining a monetary union allows a country to enjoy completely free trade in goods and services, free movement of labor (workers can travel across country borders for work), and free movement of capital (firms can build factories and invest in assets across borders). In addition, a shared currency removes all remaining possible barriers to trade, as countries can no longer intervene in their forex markets to give their exporters or importers an unfair advantage over foreign competition.

Monetary unions are therefore the highest level of economic integration possible without countries abandoning their national governments' abilities to control fiscal policy.

Advantages of monetary union include:

- Increased allocative efficiency across member countries (trade creation)
- Increased competition
- Larger export markets
- More product variety and lower prices for consumers
- More employment opportunities for workers who can travel across national borders for work
- Greater access to labor and capital for firms wishing to increase output
- Economies of scale for firms producing for a larger market

Despite the advantages, monetary unions do have their disadvantages for both member and non-member countries.

- Trade diversion may result as member countries place common external tariffs on

non-member states

- Loss of economic sovereignty as national industries are increasingly owned by foreign businesses and investors
- Downward pressure on wages as increased size of the labor force increases competition in labor markets
- Loss of monetary sovereignty as countries' central banks are disbanded in favor of the currency union's central bank
- Loss of ability to manage the currency's exchange rate as foreign exchange policy is determined by the currency union's central bank

The main disadvantages arise from the loss of economic sovereignty. Of course, some of these disadvantages can be thought of as secondary to the economic advantages of increased integration and greater efficiency in resource allocation. For example, the nationality of a factory's owner should be less important from an economic perspective than the efficiency with which the factory produces its output, the employment opportunities it creates, and the contribution to overall output it provides.

Beyond monetary unions, the highest level of economic integration is **complete economic integration**, which essentially describes a sovereign country's national economy. For example, the 50 states of the United States of America are completely economically integrated. The United States enjoys:

- Completely free trade in all goods and services
- Free movement of capital and labor across states lines
- Common external tariffs set by the US federal government
- A single currency managed by the US Federal Reserve Bank
- A single federal government determining fiscal policy for all 50 states, collecting taxes and spending on public goods and transfers benefiting the entire country.

The essential difference between the United States of America and the 18 Eurozone countries is that the US has a single fiscal authority that can tax all US citizens and provide goods and services to all US citizens. In contrast, a citizen of one Eurozone country (say, Greece) pays taxes to her own government and is eligible to receive benefits from her own government, but does NOT pay taxes to other Eurozone countries' governments and does not receive benefits from those governments. A Greek citizen cannot collect a pension in her old age from the French government.

Eurozone countries' governments manage their own national budgets and can run surpluses or deficits based on their own fiscal priorities. Individual American states, on the other hand, must balance their budgets each year because they cannot issue bonds to finance deficits as national governments can.

The fiscal sovereignty of countries in a monetary union brings up one last disadvantage of such agreements: if individual countries are fiscally irresponsible (e.g. run persistent, large budget deficits), the very legitimacy of the currency union could be undermined, affecting the economies of all other member countries. By joining a currency union with 18 other

countries, a new Eurozone economy is putting faith in the governments of its 18 currency partners to maintain relatively balanced budgets and not drive up their national debts to a level that threatens the stability of the currency.

The Eurozone financial crisis of 2011 demonstrated the risks to countries that had joined the Eurozone, when five member countries (Portugal, Italy, Ireland, Greece, and Spain) ran up deficits that spooked international investors and threatened the stability of the entire monetary union.

3.5 Terms of Trade (HL only)

Meaning and measurement of terms of trade

- Explain the meaning of the terms of trade.
- Explain how the terms of trade are measured.
- Distinguish between an improvement and a deterioration in the terms of trade.
- Calculate the terms of trade using the equation: Index of average export prices/index of average import prices x 100.

In considering a country's trade relationship with other countries, it is useful to have a measurement of the value of a country's exports relative to its imports. **Terms of trade (ToT)** provide just such a measurement. A country's ToT represents an index of its export prices relative to an index of its import prices.

$$\text{Terms of trade} = \frac{\text{Index of average export prices}}{\text{Index of average import prices}} \times 100$$

The value of a country's ToT tells us how many units of imports a single unit of exports can purchase. A country experiences an **improvement in its terms of trade** when its export prices increase relative to its import prices, and a **deterioration in its terms of trade** when its export prices decline relative to import prices.

Recall from our chapter on the macroeconomic objectives that inflation is measured using a price index consisting of all the goods bought by the typical consumer in a country. Terms of trade is measured using two price indexes: one including all the goods exported by a country and one including the goods imported.

For example, consider the simplified price indexes of Country Z's exports and imports below:

Country Z's exports	2018 price (\$)	2019 price (\$)	Country Z's imports	2018 price (\$)	2019 price (\$)
Cotton	20	22	Smartphones	200	240
Wheat	15	14	Cars	700	720
Rice	25	30	Sunglasses	100	90
2018 export price index:			2018 import price index:		
$\frac{20+15+25}{20+15+25} \times 100$ $= \frac{50}{50} \times 100$			$\frac{200+700+100}{200+700+100} \times 100$ $= \frac{1000}{1000} \times 100$		

= 100	= 100
2019 export price index: $\frac{22+14+30}{50} \times 100$ $= \frac{66}{50} \times 100$ = 132	2019 import price index: $\frac{240+720+90}{1000} \times 100$ $= \frac{1050}{1000} \times 100$ = 105

The simple, three-good examples above have provided us with values for Country Z's export and import price indexes for two years. We can summarize these values in a new table.

Export price index		Import price index	
2018	100	2018	100
2019	132	2019	105
2018 terms of trade: $\frac{100}{100} \times 100$ = 100		2019 terms of trade: $\frac{132}{105} \times 100$ = 125.7	

Let's analyze the results from our calculations above. In the first table, we calculated the prices indexes for exports and imports using 2018 as a base year. Of course, the base year price indexes are both 100, so the country's terms of trade for 2018 is 100.

In 2019 we saw the prices of some goods rise and some fall, but overall both the index of export prices rose (from 100 to 132) and the index of import prices rose (from 100 to 105).

Using the 2019 price indexes, we then calculated Country Z's terms of trade for 2019 by dividing the export price index by the import price index and multiplying the result by 100, giving Country Z a new ToT of 125.7. This is higher than 2018's value, indicating that Country Z's ToT **improved**, which means the average value of its exports increased by more than the average value of its imports.

Let's look at one more example. In this case we'll skip the first step in calculating ToT and look only at the table of price indexes for imports and exports, this time for 2019 and 2020.

Export price index		Import price index	
2019	132	2019	105
2020	134	2020	112
2019 terms of trade:		2020 terms of trade:	

$\frac{132}{105} \times 100$ = 125.7	$\frac{134}{112} \times 100$ = 119.6
---	---

In 2020 both the export price index and the import price index rose again, indicating that both the goods exported by Country Z and the goods imported went up in price. However, the terms of trade fell, or **deteriorated**, from 125.7 to 119.6. In other words, the relative price of exports to imports decreased between 2018 and 2019. Country Z's export prices rose, but not by as much as its import prices.

Another way to interpret the ToT is as the number of imports a single unit of exports could buy:

- In 2019 a single unit of Country Z's exports could buy $\frac{125.7}{100}$, or 1.257 units of imports.
- In 2020 a single unit of exports could buy only $\frac{119.6}{100}$, or 1.196 units of imports.

An **improvement in its terms of trade** means a country can afford to buy more imports for every unit of its exports it sells to the rest of the world, while a **deterioration in its terms of trade** means a country can afford to buy fewer imports for every unit of its exports.

Causes of changes in the terms of trade

- Explain that the terms of trade may change in the short term due to changes in demand conditions for exports and imports, changes in global supply of key inputs (such as oil), changes in relative inflation rates and changes in relative exchange rates.
- Explain that the terms of trade may change in the long term due to changes in world income levels, changes in productivity within the country and technological developments.

Short-term changes in terms of trade

A country's terms of trade will change when there is a change in the prices of its exports or of its imports, a change in relative inflation rates, or a change in the exchange rate.

Changes in export demand:

- If demand for a country's exports increases, ceteris paribus, its terms of trade will improve.
- If demand for a country's exports decreases, terms of trade deteriorate.

Changes in export supply:

- An increase in the supply of goods a country produces for export will reduce their prices and lead to a deterioration in the terms of trade.
- A reduction in supply of goods a country produces for exports will cause export prices to rise and lead to an improvement in terms of trade.

Changes in relative inflation rates:

- An increase in domestic inflation relative to inflation abroad will cause a country's terms of trade to improve as its goods increase in price relative to those produced in other countries.
- Deflation or disinflation (or higher inflation in other countries) will cause a country's terms of trade to deteriorate.

Changes in relative exchange rates:

- An appreciation of a country's currency will cause its terms of trade to improve as its exports become more expensive and imports become cheaper.
- A depreciation of the currency will cause a deterioration in the terms of trade as imports become more expensive and exports become cheaper.

Short-term factors that cause a change in terms of trade are summarized in the table below.

Factors that cause an improvement in the ToT	Factors that cause a deterioration in the ToT
Increase in demand for exports	Decrease in demand for exports
Decrease in supply of exports	Increase in supply of exports
Increase in domestic inflation (relative to trading partners)	Decrease in domestic inflation (relative to trading partners)
Appreciation of currency	Depreciation of currency

Long-term changes in terms of trade

Over the long term a country's terms of change can change due to changes in world income levels and improvements in productivity and technology.

If the technology used to produce a country's exports improves, productivity will increase and the relative prices of exports will fall, leading to a deterioration in terms of trade. On the other hand, if technology and productivity in foreign countries improve, imports will get cheaper and terms of trade will improve.

In reality, changes in technology and productivity are going on simultaneously around the world, leading to constantly changing terms of trade among countries participating in the global economy.

Consequences of changes in the terms of trade

- Explain how changes in the terms of trade in the long term may result in a global redistribution of income.
- Examine the effects of changes in the terms of trade on a country's current account, using the concepts of price elasticity of demand for exports and imports.
- Examine the impacts of short-term fluctuations and long-term deterioration in the terms of trade of economically less developed countries that specialize in primary commodities, using the concepts of price elasticity of demand and supply for primary products and income elasticity of demand.

Terms of trade and global income distribution

Countries that experience a long-term improvement in their terms of trade can expect to enjoy rising incomes over the long-term, as fewer units of their exports must be sold in order to buy a single unit of imports.

On the other hand, a long-term deterioration means that more exports must be sold in order to buy a single unit of imports. Countries where the ToT deteriorates over the long-term can therefore expect to enjoy a smaller share of global income relative to those where ToT improve.

Terms of trade and the current account balance

When terms of trade improve, the effect on the current account balance in the balance of payments depends on the price elasticity of demand for exports and imports

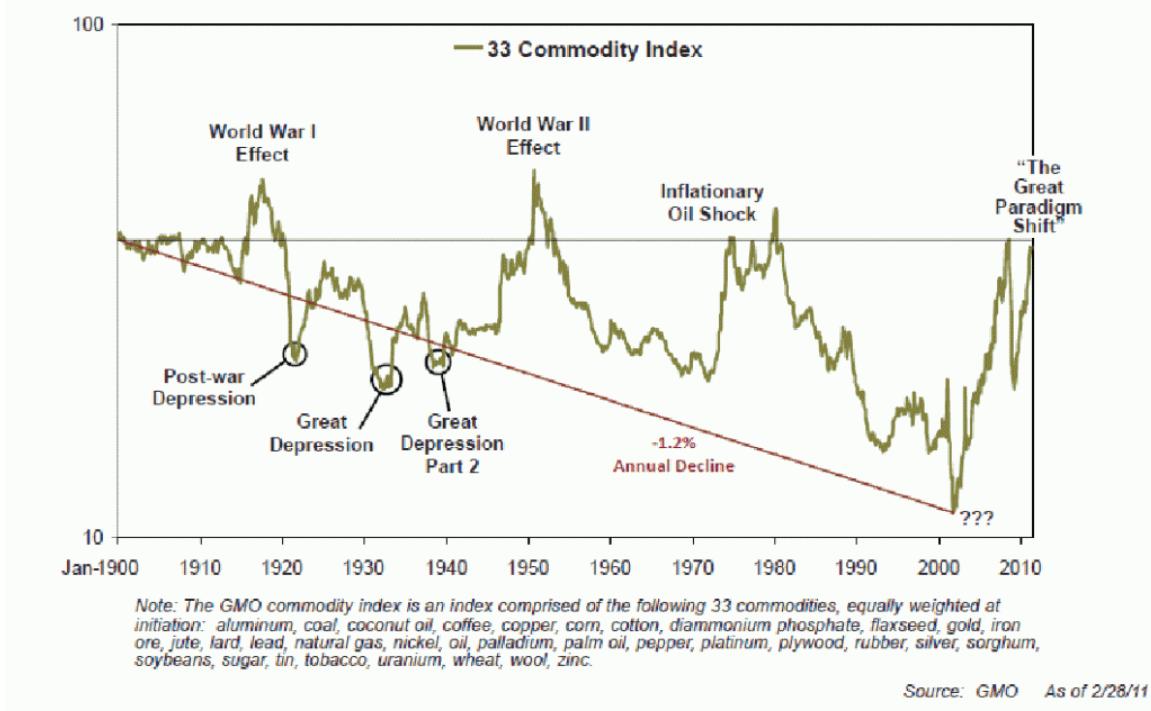
- **If demand for exports and imports is inelastic**, an improvement in the terms of trade will cause the current account balance to **move towards surplus**, since the higher price of exports will cause foreign spending on exports to increase and lower price of imports will cause spending on imports to decrease.
- **If demand for exports and imports is elastic**, an improvement in the terms of trade will cause the current account balance to **move towards deficit**, since higher priced exports will cause spending on exports to decrease and lower priced imports will cause spending on imports to increase.

Much as was the case with changes in the exchange rate (recall the Marshall-Lerner condition), the effect of a change in a country's terms of trade on the balance of payments depends on the price elasticity of demand for exports and imports.

Implications of terms of trade for commodity-dependent economies

Many of the less economically developed countries in the world specialize in the production of primary commodities. The chart below shows how the price of commodities fluctuated over the 20th and early 21st centuries.

GMO Commodity Index: The Great Paradigm Shift



Notice that with a few exceptions, the general trend has been falling commodity prices. For countries that specialize in primary commodities, this means a long-term deterioration in their terms of trade. More commodities must be produced and exported in order to continue to afford much needed imports, such as capital goods and technology, medical equipment, communication technologies, and other items essential for economic development.

For poor, developing countries, an over-dependence of the production and export of primary commodities can mean a long-run decline in living standards. In order to avoid this trap, countries should develop sectors of their economy that do not depend on non-renewable primary commodities such as minerals, energy resources, or agriculture. Both supply of and demand for these commodities are **relatively inelastic**, meaning their prices will rise and fall more than those of manufactured goods and services.

Due to the short-run volatility in commodity prices and the long-run decline represented in the chart above, an overdependence on such goods poses an obstacle to economic development for poor countries.

In our next chapter we will learn more about economic development and how poor countries can best hope to escape the poverty traps that keep them poor.

4.1 Economic Development

Economic growth and economic development

- Distinguish between economic growth and economic development.
- Explain the multidimensional nature of economic development in terms of reducing widespread poverty, raising living standards, reducing income inequalities and increasing employment opportunities.
- Explain that the most important sources of economic growth in economically less developed countries include increases in quantities of physical capital and human capital, the development and use of new technologies that are appropriate to the conditions of the economically less developed countries, and institutional changes.
- Discuss the relationship between economic growth and economic development, noting that some limited economic development is possible in the absence of economic growth, but that over the long term economic growth is usually necessary for economic development (however, it should be understood that under certain circumstances economic growth may not lead to economic development).

Throughout our study of macroeconomics we focused on the importance of Gross Domestic Product to a country's economic well-being. GDP, or the total output of goods and services produced within a nation in a period of time, represents a country's national income. Increasing GDP, therefore, generally means higher incomes for a nation's households and better living standards (although not always).

In the final section of this course we will go beyond GDP and examine the factors that contribute to not just economic growth, but economic development. **Economic development** is the improvement in the quality of life in a country, represented by rising incomes, better health, improved education, human rights, gender, racial, and religious equality, and other indicators of societal well-being.

In other words, economic development is a broader measure of welfare than economic growth, which only considers the output or income of a country, ignoring other factors that are clearly important to societal well-being.

Economic growth is a key factor in achieving economic development. Rising incomes allow people to afford a better quality of life, and usually mean better access to healthcare, better education, and overall improved living standards. However, development requires more than just higher incomes, including:

- The reduction of widespread poverty
- Reduced income inequality
- Increasing employment opportunities

Sources of economic development

Some people look at the world and believe there are two types of countries: "developed

countries” and “developing countries.” Development economists, on the other hand, understand that every country is developing and that the path towards economic development is long and never really ends. Steps can always be taken to improve the quality of life of a country’s people, even in those countries where living standards are already very high.

The primary sources of economic development lie in improvements in **human capital**, the development and adoption of **new technologies**, and **institutional changes**.

- Human capital is improved through increased quality and access to education for everyone in society and improved health, both of which increase labor productivity and lead to a happier, healthier, more educated, and higher paid population.
- New technologies appropriate to the conditions of the economically less developed countries mean productivity and efficiency increase and the country experiences both economic growth and development. This does not mean slightly faster smartphones every year, but may instead involve improvements in agriculture and manufacturing technologies, telecommunications and sanitation infrastructures, and other areas of technology that those in more economically less developed countries may take for granted.
- Institutional change can be driver of economic development. Institutions such as a country’s banking system, its legal system, the ownership and structure of its corporations, the structure of its political system, and even social institutions like religious organizations and the role of women and minority groups can all impact the level and rate of economic development in a country. Institutional reforms, such as implementing a clear legal framework for intellectual property rights, can accelerate the pace of economic growth and development and lead to meaningful improvements in people’s living standards over time.

Economic growth and economic development

We’ve all been asked to consider the question “what came first, the chicken or the egg?” In pondering the answer to this question, consider another question, one that is relevant to our current study of economics: “What comes first, economic growth or economic development?”

To some extent, economic development can happen in the absence of economic growth. For instance, in a country in which there is a huge gap between rich and poor, but which has a reasonably high per capita income, improvements in development can be achieved without further increasing the average income level. Greater equity can be achieved through redistributing the income that’s already there in a way that improves the lives of many without reducing the opportunities for all.

In other ways living standards can be improved without first pursuing sustained increases in output. Social justice, human rights, gender and racial equality, and democratization are all steps a country can take that will improve the lives of its people without requiring an increase in income or output.

However, over the long run, there is a limit to how “developed” a nation can become if its people remain stuck at a low income level. Economic growth, or increasing the average income and output in a country, is required to achieve a high level of human development. Without the resources and the financial means, people’s living standards will be limited due to a lack of access to both the public and private goods that are required for a safe, healthy, and rewarding life.

In certain circumstances, of course, economic growth may NOT lead to economic development. If rising incomes are enjoyed only by a privileged few in society, while the majority of a country’s population remains stuck in poverty, then economic growth is not leading to development. Furthermore, if rising output is accompanied by a deterioration of environmental or human health, then all the money in the world won’t make a country better off if its ecosystem is dying in the process.

The **externalities** and **inequality** that sometimes accompany economic growth can pose an obstacle for economic development. In order for growth to lead to development, it must be **sustainable**, which means gains today must not come at the expense of those who will live tomorrow. Sustainable economic growth is achieved when output and incomes increase in a way that does not jeopardize the future health of the environment or of humans.

Common characteristics of and diversity among economically less developed countries

- Explain, using examples, that economically less developed countries share certain common characteristics (noting that it is dangerous to generalize as there are many exceptions in each case), including low levels of GDP per capita, high levels of poverty, relatively large agricultural sectors, large urban informal sectors and high birth rates.
- Explain that in some countries there may be communities caught in a poverty trap (poverty cycle) where poor communities are unable to invest in physical, human and natural capital due to low or no savings; poverty is therefore transmitted from generation to generation, and there is a need for intervention to break out of the cycle.
- Explain, using examples, that economically less developed countries differ enormously from each other in terms of a variety of factors, including resource endowments, climate, history (colonial or otherwise), political systems and degree of political stability.

Note: All the bubble charts in this and subsequent chapters are created using GapMinder Tools, which harness publicly available data to produce easy to understand interactive charts. You can make your own charts similar to those seen here at <http://www.gapminder.org>

Before we dive deeper into our study of economic development, let’s get some terminology straight. As mentioned in the previous section, the idea that some countries are “developed” while others are “developing” is a false assumption. As economists, we distinguish between “economically less developed countries” and “economically more developed countries.”

- **Less economically developed countries (LEDCs)** are those that generally exhibit a lower level of human development, including relatively low life expectancy, low literacy rates, and low income.
- **More economically developed countries (MEDCs)** are those in which life expectancy, literacy rates, and average income are all relatively high.

There is no clear cut-off between what is meant by “high” or “low” in these cases, rather, a country’s level of development falls along a spectrum; some countries are clearly more developed than others, while no country is “100% developed.”

In the following section we’ll be looking at lots of charts in which different color bubbles are displayed. The map below shows which region of the world each color on these charts represents.



¹

All green bubbles represent countries in the western hemisphere, blue bubbles are African countries, and so on.

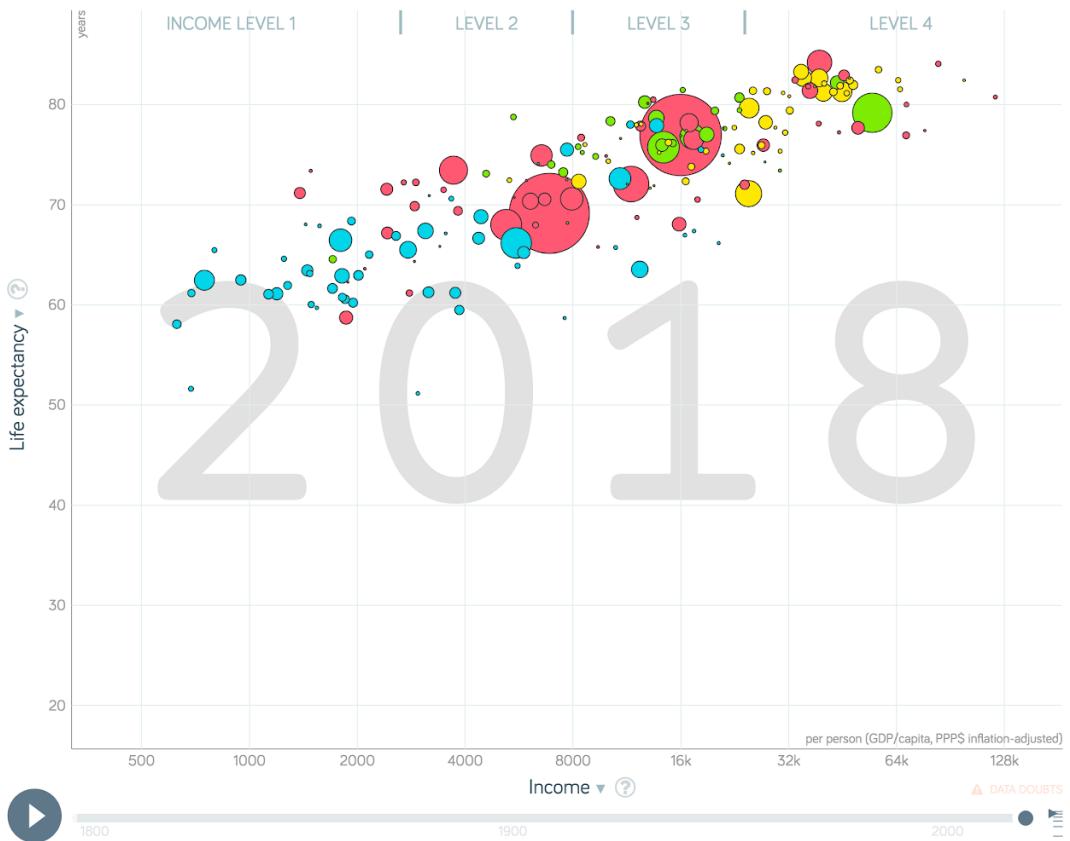
Additionally, the size of each bubble represents the country’s population. So the largest bubble in most charts will be China, since China is the world’s most populous country.

Now that you understand what you’re looking at, let’s dive into some of the common characteristics of LEDCs.

Common characteristics #1 - Short lives and low incomes

The chart below shows the life expectancy (in years) of the average citizen plotted against per capita income.

¹ Source: Gapminder.org



Observe from the chart:

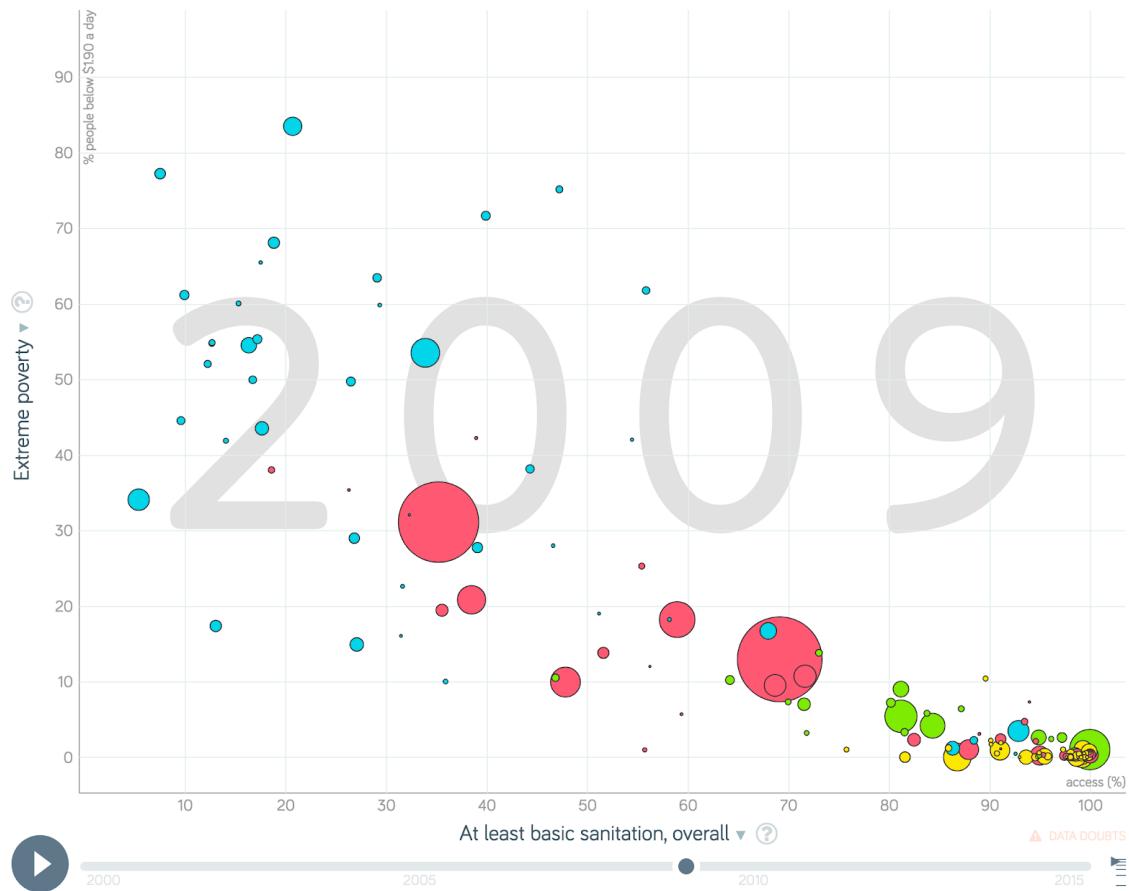
- Countries with low incomes (below \$4,000 per year) tend to have shorter life expectancies (below 70 years) while countries with higher incomes tend to have longer life expectancies (more than 80 years for most European countries).
- Most of Africa falls in the “low income/short lives” sector of the graph, with many countries averaging less than \$4000 per year and only around 60 years of life.
- Most European countries (yellow) enjoy higher incomes (more than \$32,000 per year) and longer lives (around 80+ years).
- China and India (the two big red bubbles) are right in the middle, with relatively long lives (70 years for India and around 76 for China) and medium income levels (\$7,000 for India and \$16,000 for China).

Clearly, there is a correlation between income and life expectancy. People in richer countries live longer lives, while people in poorer countries live shorter lives. We can thus conclude that shorter lives and low incomes are common characteristics of LEDCs, while longer lives and higher incomes are common characteristics of MEDCs.

Common characteristics #2 - Poverty and sanitation

Next let's look at two other indicators: poverty and sanitation. The chart below shows the percentage of people living in extreme poverty (defined as earning below \$1.90 per day) and the percentage of the population that has access to basic sanitation (including flush toilets

and piped water and sewage systems).



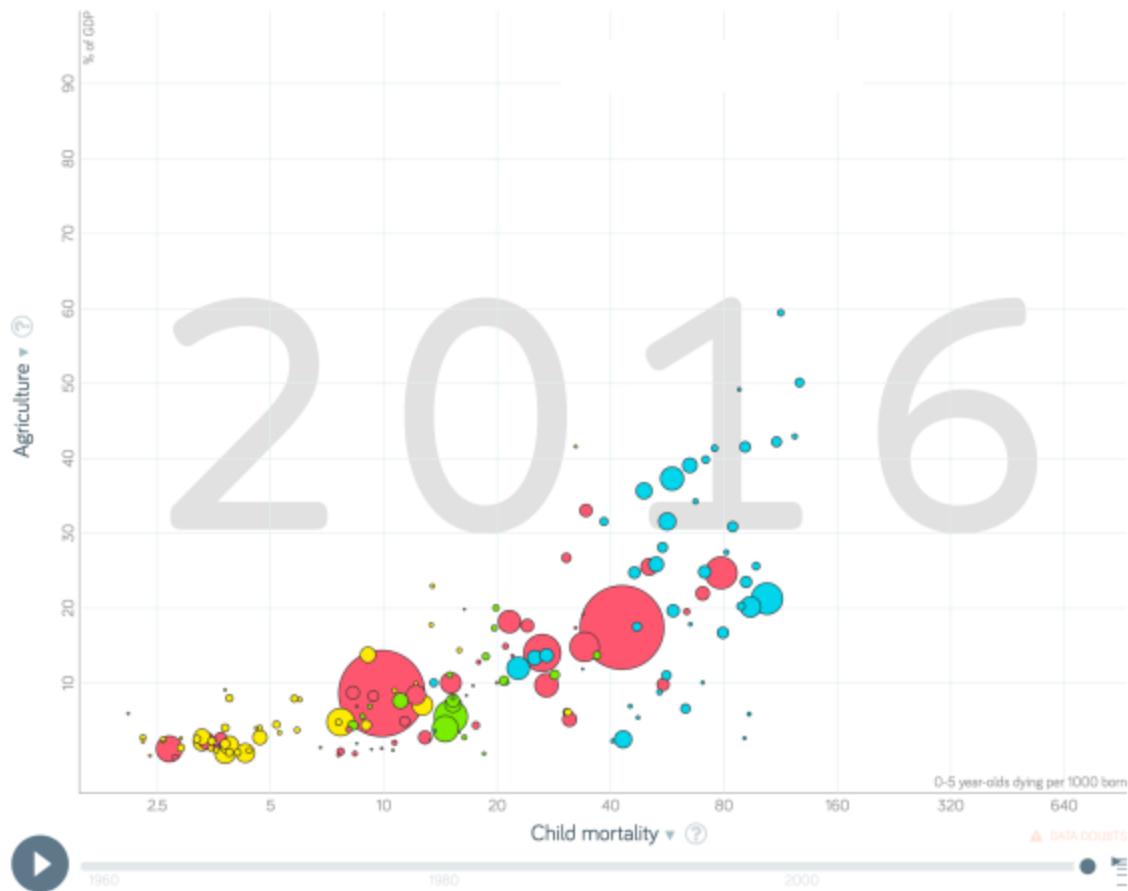
Observe from the chart:

- In countries in which fewer than 50% of the population has access to improved sanitation poverty rates are very high (mostly above 20%, in some cases as high as 80%).
- Greater access to sanitation is correlated with lower rates of poverty.
- Countries where 100% of the population have access to improved sanitation enjoy 0% poverty rates.
- Again, most of the countries suffering from poor sanitation and poverty are in Africa.

We can conclude from the chart above that lack of access to improved sanitation and high levels of poverty are common characteristics of LEDCs.

Common characteristics #3 - Agriculture and maternal health

A less obvious characteristic of LEDCs is the relative importance of agriculture to their national economies. The chart below shows the percentage of total GDP made up of the production of agricultural commodities and the child mortality rate, or the number of children per 100,000 who die before the age of 5.



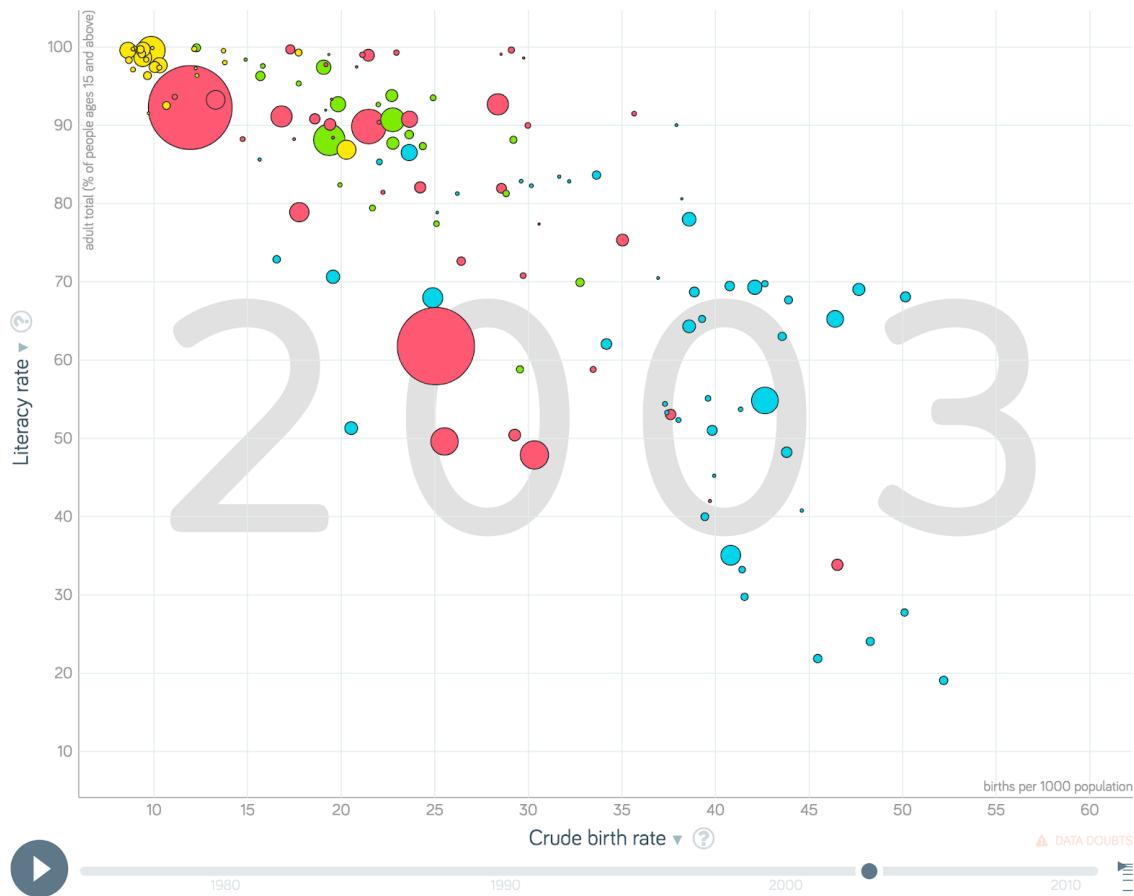
Observe from the chart:

- Countries in which agriculture makes up a small percentage of the economy (less than 10% of GDP) tend to enjoy very low child mortality rates (fewer than 5 per 1,000 births).
- Countries with relatively large agricultural sectors (more than 40% of GDP) suffer from higher child mortality (more than 40 per 1,000 births).
- In Europe agriculture represents a relatively small proportion of economic output (mostly under 10%), while in Africa agriculture represents over 20% of most countries' economies.

This begs the question: does having a relatively large agricultural sector lead to higher child mortality? The answer is almost certainly “no”. This is not an issue of causation, rather of correlation. Also, countries in which agriculture is a larger percentage of GDP likely produce far LESS food per person than countries in which agriculture is a small percentage of GDP. The difference is, in LEDCs there is little else besides agricultural commodities that gets produced, whereas in MEDCs the manufacturing and service sectors, in which workers are relatively productive and earn higher incomes than in agriculture, account for a much larger percentage of total output.

Common characteristics #4 - Poor education and high birth rates

For a final comparison of LEDCs versus MEDCs we'll consider two more factors: level of education and family size. The chart below shows the literacy rates (percentage of adults who can read) and the crude birth rate (the number of births per 1000 people in a year).



Observe from the chart:

- Countries in which everyone can read tend to have fewer children, while countries with lower literacy rate (below 70%) tend to have more children (more than 35 per 1000 in the population).
- Poor education (represented by low literacy rates) is correlated with poor access to family planning (birth control, contraceptives, etc...)

The relationship seen here may, in fact, be one of not only correlation, but to some extent, causation. It is entirely likely that the lower the rates of literacy, the less aware people are of methods for family planning and contraception. Additionally, low literacy means people are less likely to be employed in the formal sector and more likely to live a traditional, home-based lifestyle in which a woman's traditional role as child bearer and caretaker is still assumed.

LEDCs (again, many of which are in Africa) are characterized by large families and low

levels of education.

The common characteristics of LEDCs examined above are just a sampling of what sets less and more economically developed countries apart from one another. As can be seen in all the charts here, there is no obvious gap between the more and the less economically developed countries in the world today. Rather, there are countries all along the spectrum from the least to the most developed. In this regard, every country is in some way a “developing country,” that could stand to improve its performance on a myriad of economic, social, and human health indicators.

International development goals

- Outline the current status of international development goals, including the Millennium Development Goals.

In 2015 the United Nations updated the goals for global economic development it had originally created around the turn of the 21st century. The “Millennium Development Goals”, which aimed to eradicate extreme poverty by the year 2015, were replaced with the “Sustainable Development Goals”, which set a target of 2030 to achieve 17 goals relating to economic development. The goals are²:



² Source: “17 Goals to Transform the World”

<https://www.un.org/sustainabledevelopment/sustainable-development-goals/>

4.2 Measuring Economic Development

Single indicators of development

- Distinguish between GDP per capita figures and GNI per capita figures.
- Compare and contrast the GDP per capita figures and the GNI per capita figures for economically more developed countries and economically less developed countries.
- Distinguish between GDP per capita figures and GDP per capita figures at purchasing power parity (PPP).
- Compare and contrast GDP per capita figures and GDP per capita figures at purchasing power parity (PPP) exchange rates for economically more developed countries and economically less developed countries.
- Compare and contrast two health indicators for economically more developed countries and economically less developed countries.
- Compare and contrast two education indicators for economically more developed countries and economically less developed countries.

NOTE: All charts in this chapter were created using Google's Public Data explorer, a tool that allows anyone to create interactive charts using publicly available data.³

GDP vs. GNI

Throughout our study of macroeconomics we focused on gross domestic product (GDP) as a measurement of national income. Recall that GDP measures the total value of the output produced within a country in a year.

An alternative indicator of national income is **gross national income (GNI)**. GNI measures the total value of output produced by the factors of production owned by a country's citizens, regardless of where in the world the factors of production are located. For example, while China's GDP includes the value of everything produced inside of China, including goods produced by foreign owned firms, its GNP would exclude the value of goods produced by foreign-owned firms in China, but include the value of output produced by Chinese firms in other countries. Since there are more foreign firms operating in China than there are Chinese firms operating outside of China, the country's GNP is about \$300 billion less than its GDP⁴.

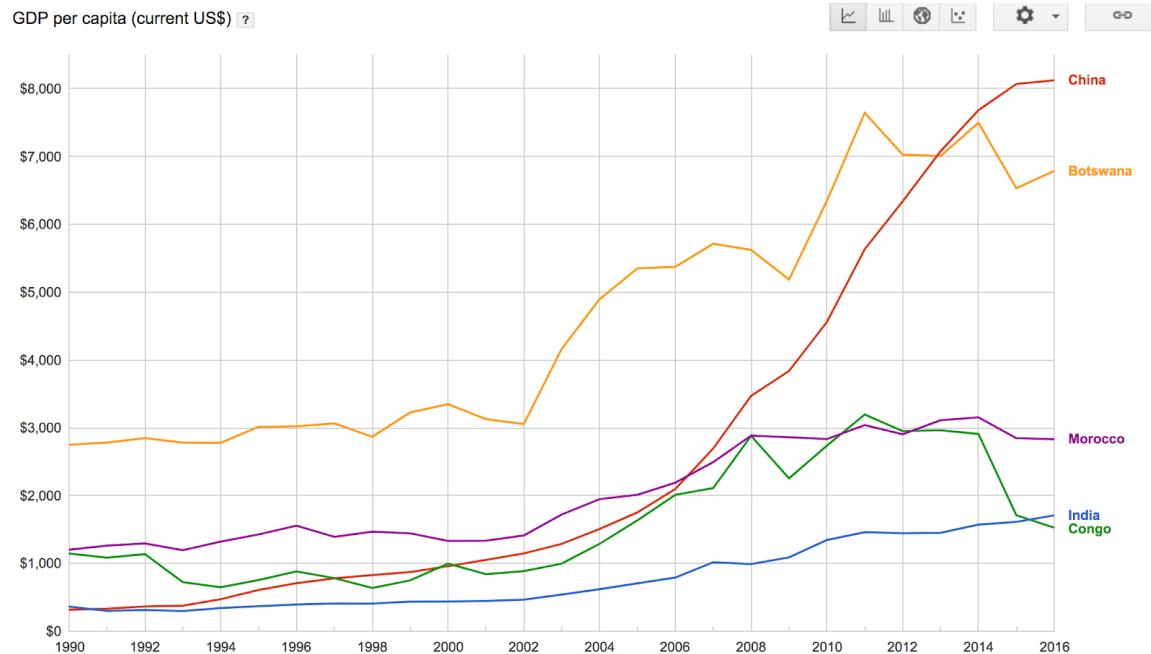
For measuring economic development, GDP tends to be the preferred indicator of income, since for the people living in a country, what matters most is the income earned within the borders of that country, regardless whose nationals own the factors of production responsible.

³ <https://www.google.com/publicdata/directory?hl=en>

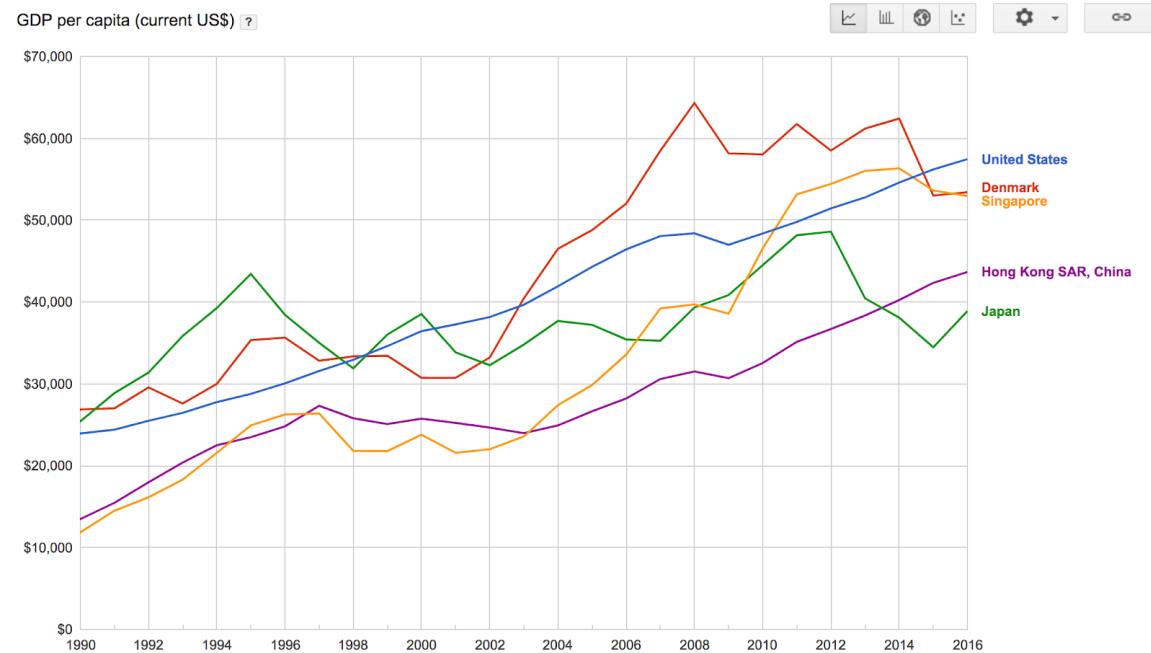
⁴ Source: Investopedia: "What is the functional difference between GDP and GNP?"

Per capita income in LEDCs and MEDCs

By dividing a country's GDP by its population, we can determine and compare the **per capita GDP** figures for LEDCs. In the chart below, we can see the per capita GDPs of five LEDCs from 1990 to 2016.



To compare, let's look at the per capita GDPs of five MEDCs over the same period.



Notice that the per capita GDPs of the five MEDCs are much higher than those of the

LEDCs. In fact, Singapore, the poorest of the MEDCs in 1990, still produced \$4,000 more output per person (\$12,000) than China did in 2016 (\$8,000).

The average of the MEDCs in 2016 is roughly \$50,000 per person per year. Compared to the average income among the LEDCs of around \$5,000, the average citizen of the MEDCs represented here earns about ten times as much each year than the average citizen of the LEDCs.

Of course, these are only 10 of the over 200 countries in the world, and a broader investigation would reveal that there are countries all along the income spectrum from the poorest to the richest.

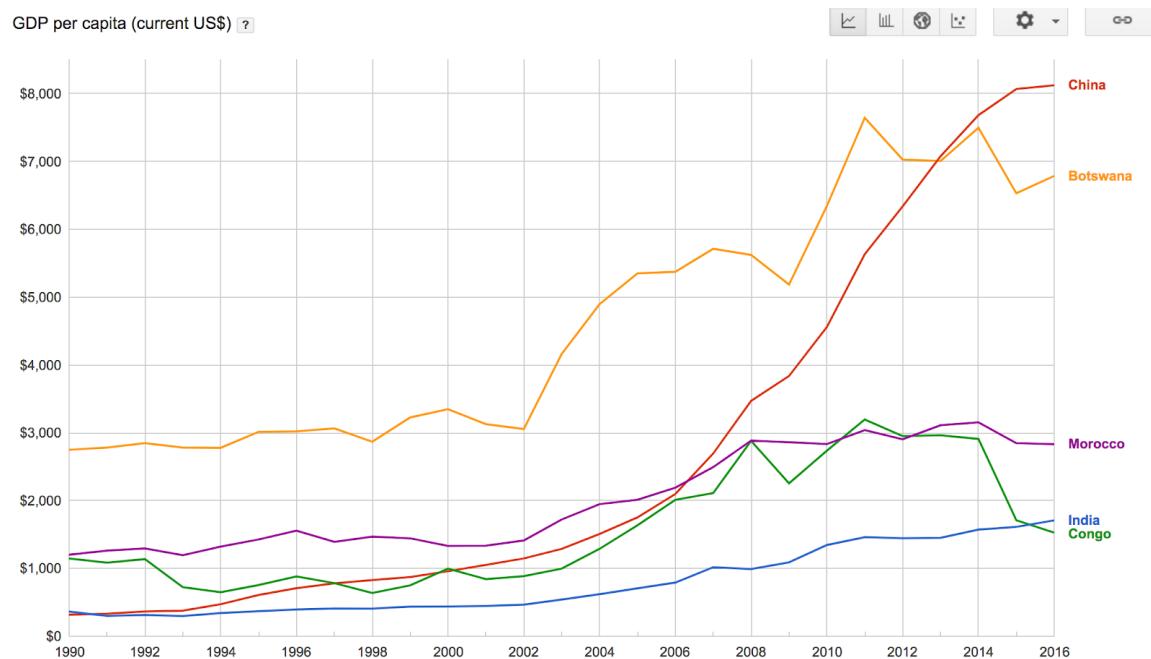
Purchasing Power Parity GDP

The per capita GDP figures in the charts above are found simply by dividing the dollar value of the countries' output by their populations. There is one major shortcoming in using per capita GDP to compare the average incomes across countries - the cost of living varies widely across different countries, therefore the dollar income may not accurately represent how "rich" or "poor" the average person feels.

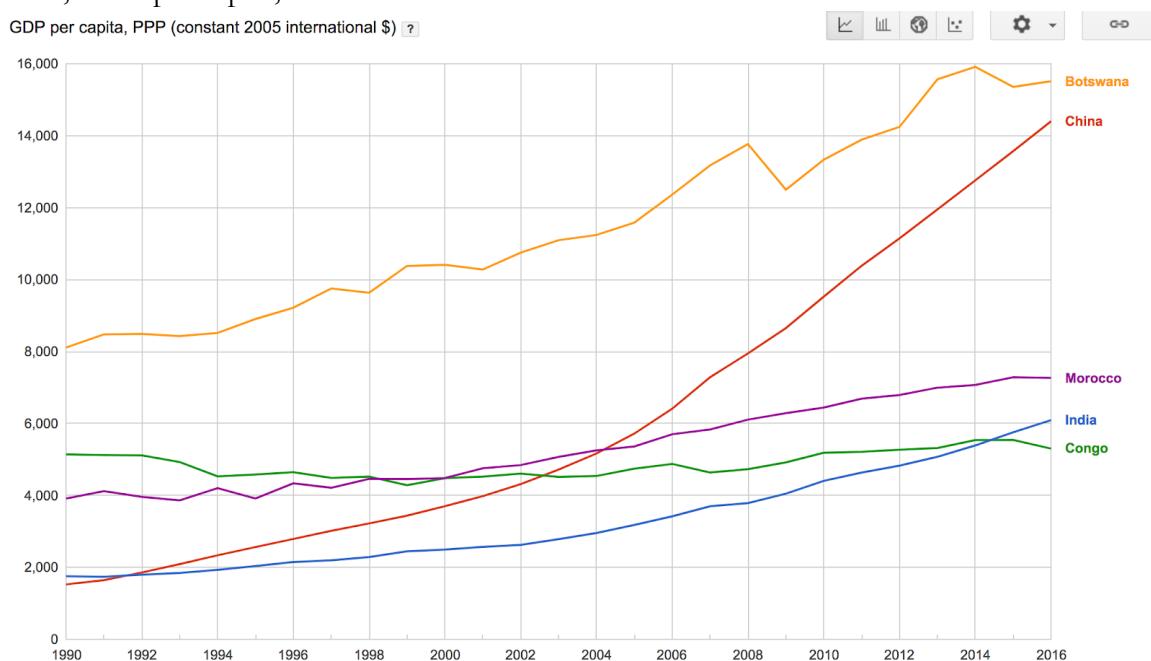
To compensate for this, economists measure what is called **purchasing power parity GDP and GDP per capita (GDP per capita, PPP)**. GDP per capita, PPP adjusts the dollar value of average income up or down depending on whether the cost of living in the country is relatively high (in which case, the GDP per capita figure is adjusted down) or relatively low (GDP per capita is adjusted up).

Let's look at those LEDCs' per capita GDPs again, but this time we'll compare the dollar value of the average income to the PPP value of average income.

First, simple GDP per capita:



Next, GDP per capita, PPP:



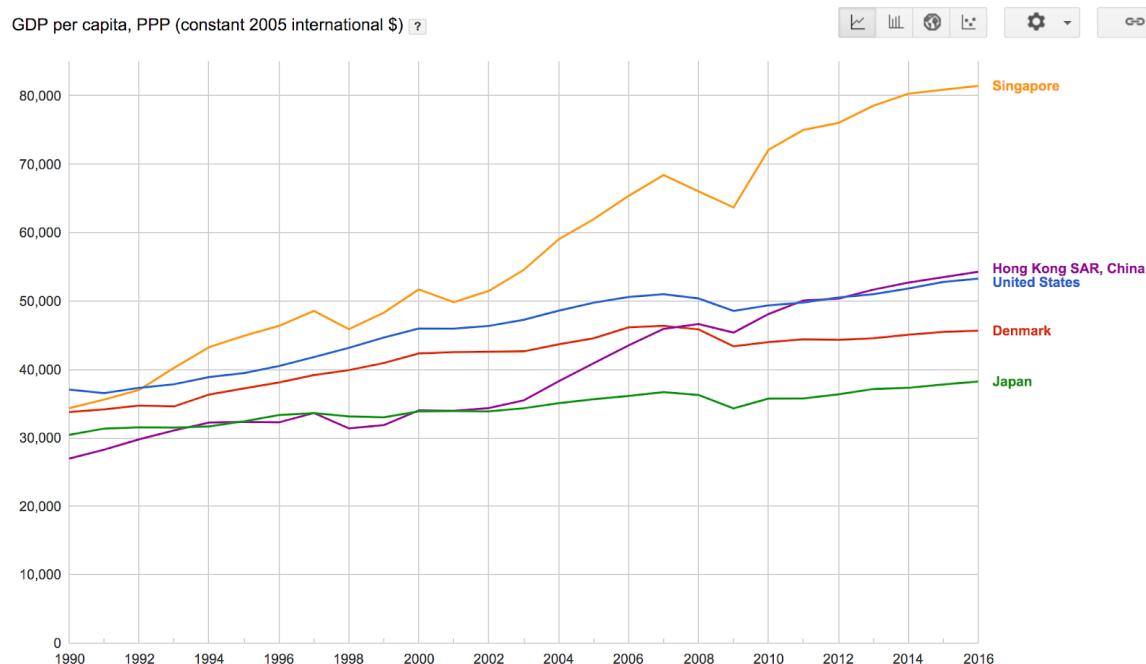
Let's compare 2016's GDP per capita to GDP per capita, PPC for each of the countries.

Country	GDP per capita	GDP per capita, PPP	Difference
China	\$8,100	\$14,300	+\$6,200

Botswana	\$6,800	\$15,400	+\$8,600
Morocco	\$2,850	\$7,200	+\$4,350
India	\$1,700	\$6,000	+\$4,300
Congo	\$1,500	\$5,300	+\$3,800

In all five countries, GDP per capita, PPP is significantly higher than the dollar value of GDP per capita. The reason for this? Living in poor countries is quite a bit cheaper than living in richer countries; so that \$2,850 of income earned by the typical Moroccan goes a lot further in Morocco than it would, say, in the United States.

To compare, let's look at the per capita GDP, PPPs in the MEDCs and compare them to the per capita GDPs for those five countries.



Country	GDP per capita	GDP per capita, PPP	Difference
United States	\$57,000	\$53,000	-\$4,000
Denmark	\$54,000	\$46,000	-\$8,000
Singapore	\$53,000	\$81,000	+\$28,000
Hong Kong	\$43,000	\$55,000	+\$12,000

Japan	\$39,000	\$39,000	+\$0
-------	----------	----------	------

Among the MEDCs, we can conclude that a dollar of income goes further in Singapore and Hong Kong than it does in the United States or Denmark. The GDP per capita, PPP is higher in countries where the cost of living is relatively low and lower in countries where living costs are relatively high.

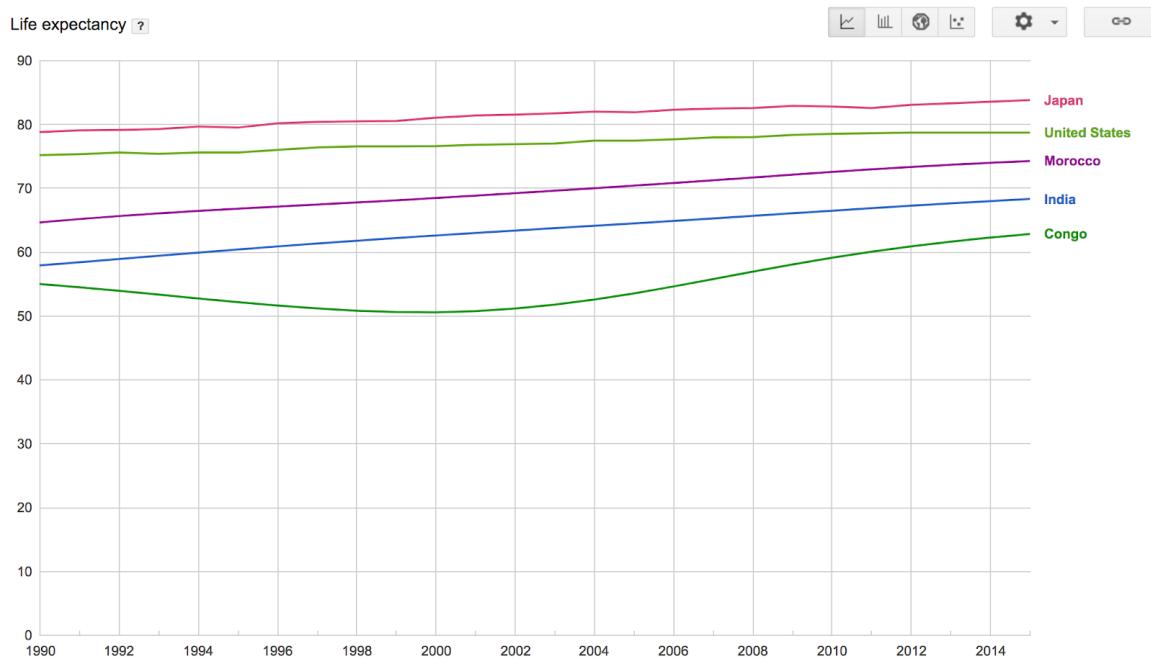
By looking at GDP per capita, PPP instead of the dollar value of GDP per capita, the gap between the richest and the poorest countries looks much narrower, since lower incomes go a lot further in poor countries than they would in richer countries thanks to the lower costs of living.

Alternative indicators - Health

As we have said, economic development means more than rising incomes. A decent standard of living also requires access to healthcare, which in turn leads to a longer, happier, more productive life.

LEDCs and MEDCs differ not only in per capita incomes, but in health as well. Two indicators we can use to compare the health are life expectancy at birth (how long the typical person will live) and the child mortality rate (# of children per 1,000 who will die before the age of five).

First, let's compare the life expectancies in five countries:

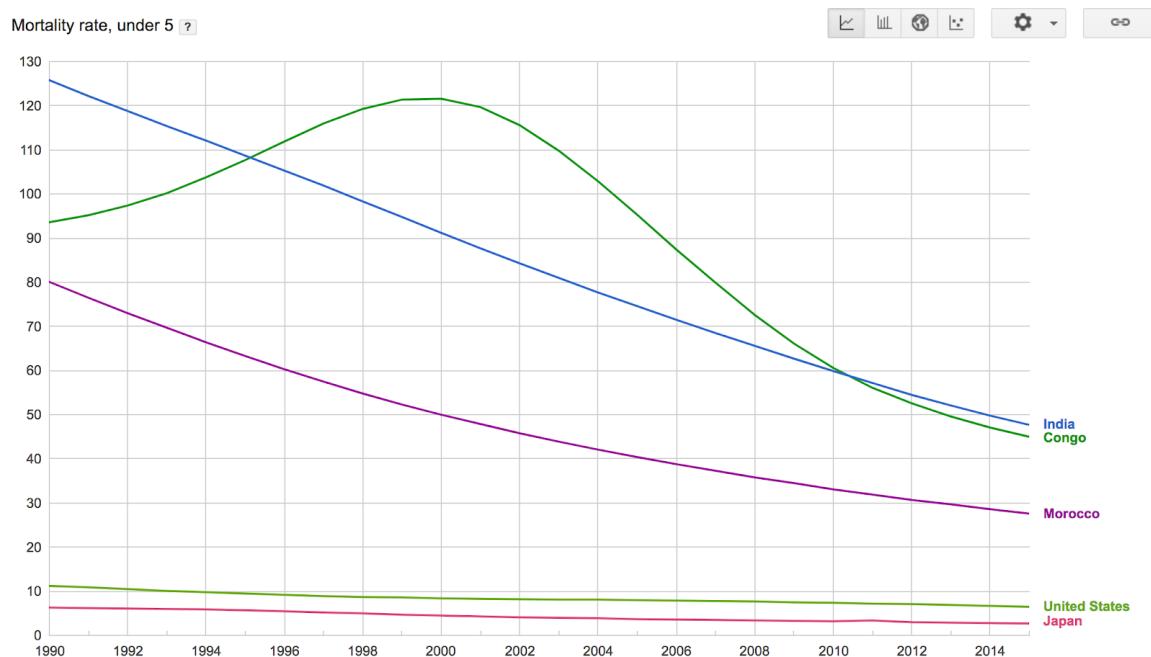


While all five countries have experienced improvements in life expectancy since 1990, there is a strong correlation between incomes and lifespans. The poorest of the five countries,

Congo, has the shortest life expectancy, at just 63 years, while India and Morocco, which are also LEDCs, likewise experience relatively short lives compared to the higher income countries.

However, at the top end, it is NOT the richest country that has the longest lives. Japanese citizens, whose per capita income is almost \$20,000 lower than the Americans', enjoy significantly longer lives than Americans (84 years versus 79 years). This points to the fact that income level is not the only indicator of economic development, and that rich countries like the United States can still experience development by allocating more resources towards health and other life-quality improving areas.

Next, let's look at child mortality for the same five countries:



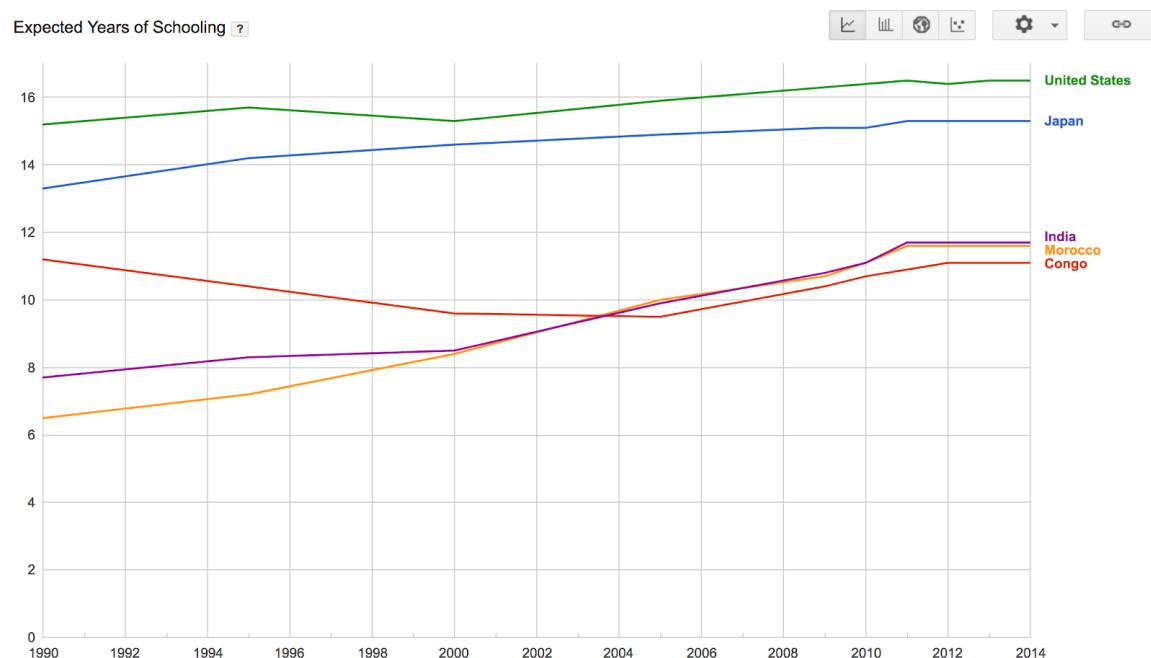
Here the differences between the MEDCs and the LEDCs are more striking. In Japan the odds of a child dying before the age of five are almost zero: just 3 per 1,000 (or 0.3 percent) children will die before their fifth birthday. In the US the number is more than double, at 8 per 1,000. However, the real gap between more and less economically developed countries can be seen when we look at the three LEDCs, where child mortality rates are 2.8% in Morocco, 4.5% in Congo and nearly 5% in India.

Both shorter lives and higher child mortality are the result of the underprovision of basic healthcare and sanitation. With improved access to these **merit goods**, the populations of LEDCs would enjoy longer lives, healthier children, and improved health overall, allowing for improved quality of life, more productivity, higher incomes, and all the other benefits that come with a higher standard of living.

Alternative indicators - Education

A third indicator of economic development, along with income and health, is the level of education of a country's people. To conclude our section on individual indicators of development, we'll consider expected years of schooling and the percentage of adult females with at least some secondary (high school) education.

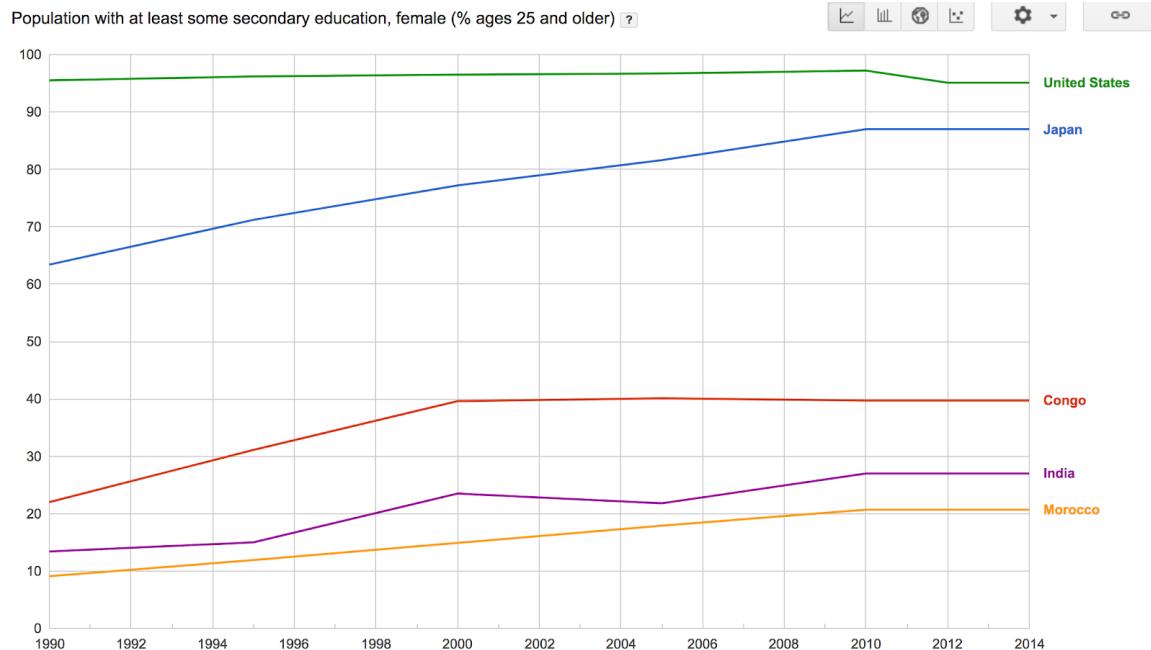
First, let's look at the number of years a child entering school can expect to complete in our five countries:



Here we see a noticeable gap between the LEDCs and the MEDCs. In Japan and the United States the typical child entering preschool or kindergarten can expect to receive between 15 and 17 years of schooling (in other words, the average person will receive at least some post-secondary education).

In the LEDCs, on the other hand, the typical child entering school can expect to receive between 11 and 12 years of schooling. The extra years of schooling (along with the higher quality and quantity of resources available in their national education systems) will give pupils in the MEDCs a better chance of finding employment in a job that pays a wage decent enough to maintain a high quality of life. The underprovision of education in LEDCs, meanwhile, will result in less productive workers, lower incomes, increased poverty, and a less productive economy as a whole.

Next, let's consider another education indicator - the percentage of females above the age of 25 who have received at least some secondary education:



Again we see a significant gap between the MEDCs and the LEDCs in this indicator, which says something not only about the level of education of the countries, but the role of women in society as well. In Japan, a society in which women fulfill a more traditional gender-specific role than in the United States, fewer than 90% of adult women have received secondary education, compared to 95% in the US.

Among the LEDCs we see a huge deficit in females' access to education. Morocco, the richest of the LEDCs, is a muslim society where women are more likely to fill a traditional gender-specific role, and as a result only 20% of adult women have received any secondary education. In India the number is 28% and in Congo 40%.

Access to education is an indicator of economic development, and not just because being educated makes individuals better off, but because an educated population is also more likely to contribute to a well-functioning, productive economy and society.

Composite indicators of development

- Explain that composite indicators include more than one measure and so are considered to be better indicators of economic development.
- Explain the measures that make up the Human Development Index (HDI).
- Compare and contrast the HDI figures for economically more developed countries and economically less developed countries.
- Explain why a country's GDP/ GNI per capita global ranking may be lower, or higher, than its HDI global ranking.

When comparing levels of economic development, it can seem tedious to determine which income, health, education, or other indicators to select in our comparisons. To simplify the

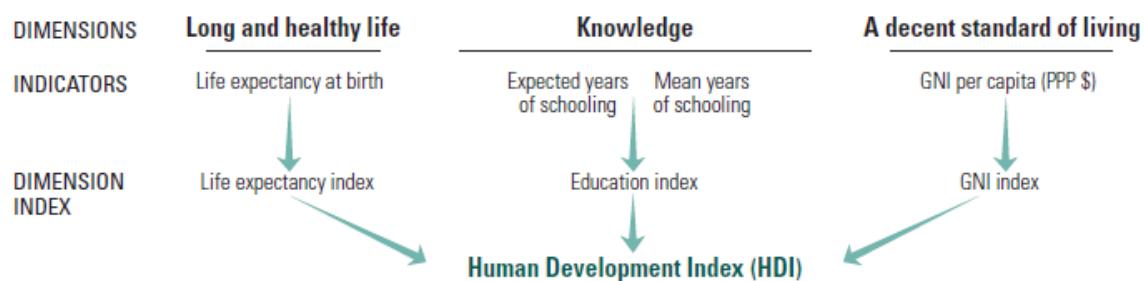
process of indicating and comparing the levels of economic development of different countries, the United Nations has developed a **composite indicator of economic development** called the **Human Development Index (HDI)**.

According to the United Nations Development Program,

“The HDI was created to emphasize that people and their capabilities should be the ultimate criteria for assessing the development of a country, not economic growth alone... The Human Development Index (HDI) is a summary measure of average achievement in key dimensions of human development: a long and healthy life, being knowledgeable and have a decent standard of living.

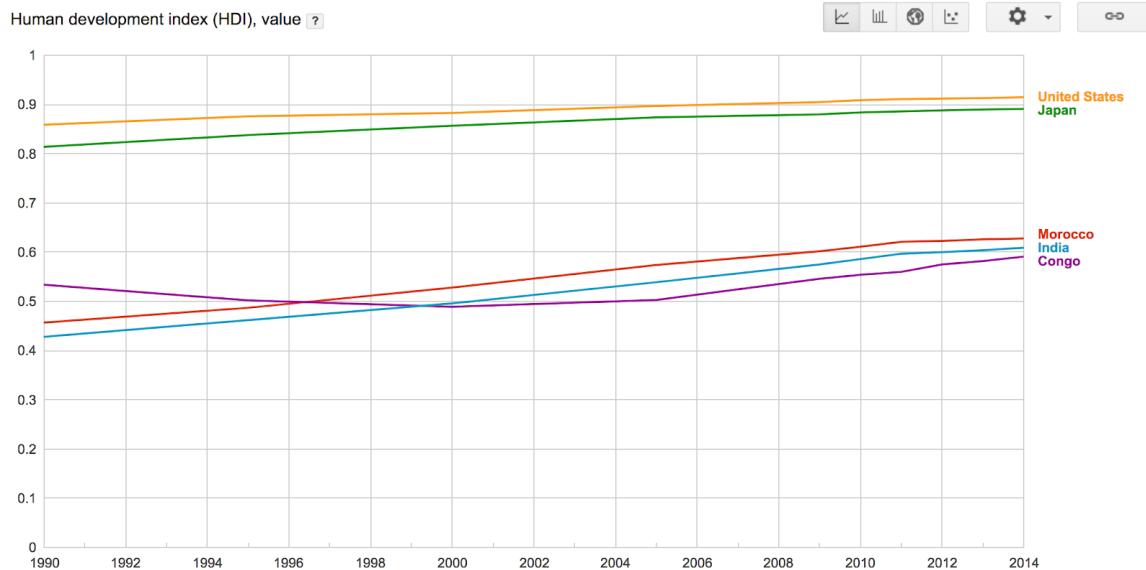
The health dimension is assessed by life expectancy at birth, the education dimension is measured by mean of years of schooling for adults aged 25 years and more and expected years of schooling for children of school entering age. The standard of living dimension is measured by gross national income per capita. The HDI uses the logarithm of income, to reflect the diminishing importance of income with increasing GNI. The scores for the three HDI dimension indices are then aggregated into a composite index using geometric mean.”

The graphic below shows how a country's HDI score is determined⁵:



Not surprisingly, MEDCs and LEDCs show a noticeable gap in their HDI scores. Let's compare the same five countries examined in the previous questions, looking now at their HDI scores (which fall within a range of 0 to 1.0).

⁵ Source: United Nations Development Program Human Development Reports - <http://hdr.undp.org/en/content/human-development-index-hdi>



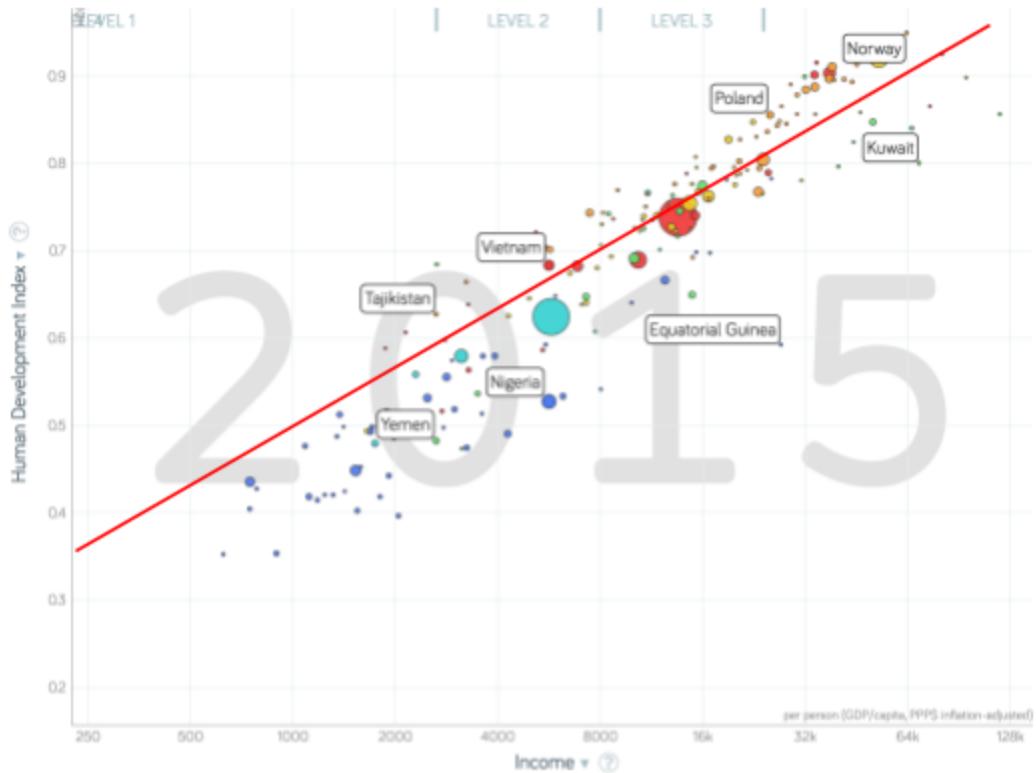
Notice from the chart:

- All five countries enjoy a higher HDI score in 2014 than they did in 1990, indicating that over the 24 year period they all enjoyed economic development.
- Congo experienced a fall in its HDI score in the 1990's and early 2000's. This was likely due to a fall in the health indicator, life expectancy at birth, due to more deaths from conflict and the HIV epidemic that swept across Sub-Saharan Africa during those years.
- The two MEDCs, the US and Japan, score better on the composite index than the three LEDCs.

Interpreting the HDI scores for these five countries we can conclude that the citizens of the US and Japan enjoy longer lives, have a higher level of education, and higher incomes than those of the three LEDCs.

HDI vs. income

While there is an obvious correlation between higher incomes and a higher HDI score (since per capita GNI is a component of the HDI score), it is not always true that a higher income guarantees a higher HDI score. In the chart below we can see the HDI scores on the vertical axis plotted against average income on the horizontal axis.



Notice from the graph:

- The red line shows the trend relationship between income and HDI. Higher incomes mean a higher HDI score, in general.
- The eight selected countries show the exceptions to the trend. Countries above the red line (Tajikistan, Vietnam, Poland, and Norway) perform better than average for their incomes, while countries below the red line (Yemen, Nigeria, Equatorial Guinea, and Kuwait) perform below average for their income levels.

It is possible for a country's HDI ranking to be higher than its global income ranking if the country has achieved gains in health and education that are disproportionate to their gains in income. This is the case for the four countries above the red line in the chart above.

On the other hand, if a country's income grows faster than improvements in health and education, its HDI ranking will be lower than its global income ranking. This is the case for the four countries below the red line.

While residents of Yemen and Tajikistan earn roughly the same income, citizens of Tajikistan are likely to be healthier and better educated than those in Yemen. The same applies to Vietnam and Nigeria, Poland and Equatorial Guinea, and Norway and Kuwait. While each of these pairs of countries have roughly the same per capita incomes, the first of each pair have made more investments in human welfare-improving merit goods like education and health, which have resulted in their HDI rankings being higher than their global income rankings.

4.3 The Role of Domestic Factors in Economic Development

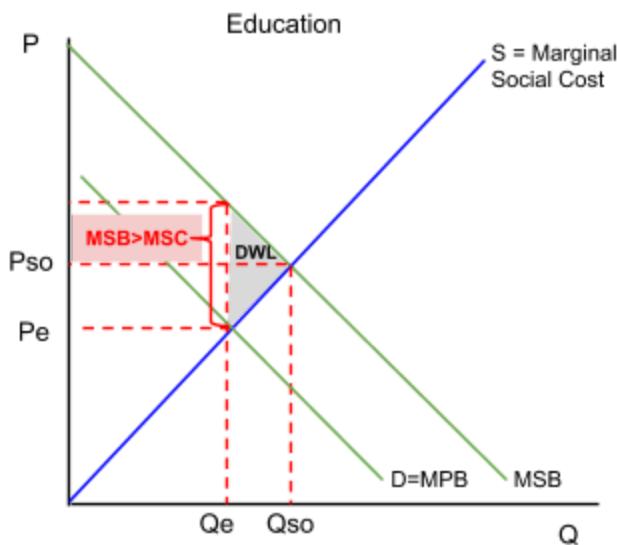
Domestic factors and economic development

- With reference to a specific developing economy, and using appropriate diagrams where relevant, examine how the following factors contribute to economic development.
 - a. Education and health
 - b. The use of appropriate technology
 - c. Access to credit and micro-credit
 - d. The empowerment of women
 - e. Income distribution

Education and health

The provision of certain **merit goods** can improve a country's level of economic development. Both education and health are goods that are underprovided by the free market, due to the **positive externalities** they provide to society as a whole.

Recall from our study of microeconomics that a positive externality can be illustrated using a marginal benefit/marginal cost diagram. Consider the market for education below, in which the consumption of education creates spillover benefits for society as a whole.



Observe from the graph:

- The supply and demand of education in the free market are determined by the marginal private benefit (MPB) and the marginal social cost (MSC). These determine the equilibrium quantity of education that will be provided by the market (Q_e)
- The marginal social benefit of education (MSB) is greater than the MPB, due to the external benefits education provides society. These include a more productive

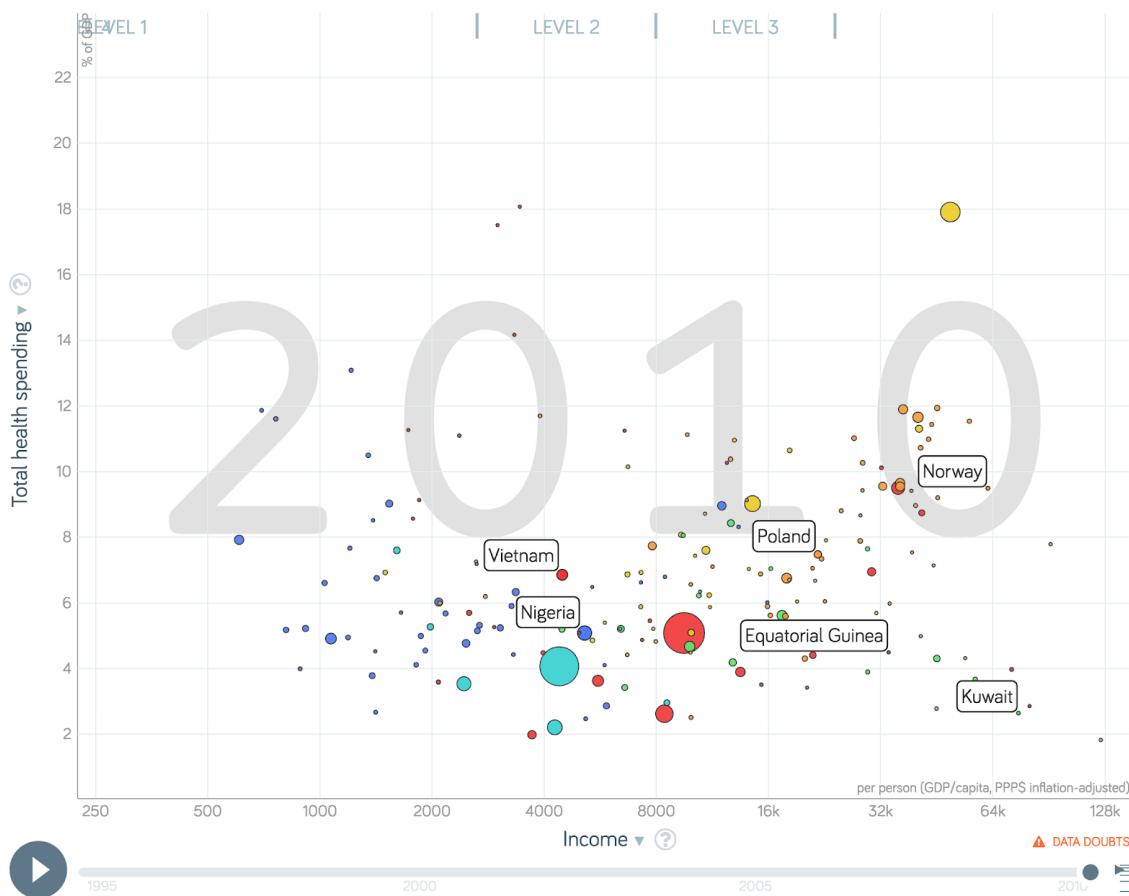
workforce, higher incomes and more taxes paid to the government, more employment, less poverty, less crime, and so on.

- The socially optimal quantity of education (Q_{so}) is greater than the free market equilibrium (Q_e). In other words, if left to the free market, not enough education will be provided to a country's people.

By investing in a country's education system, its government can expect improvements in both economic output and human development.

The same goes for health: because the benefits of healthcare are not purely private (in other words, society as a whole benefits from a healthier population), the free market quantity of healthcare will be less than the socially optimal quantity. Government support for healthcare, through a national health system, subsidies to insurance companies or to health insurance buyers, or other interventions in the market for healthcare or health insurance, will result in a healthier and more productive population, more economic output, and improvements in human development.

The chart below compares some of the countries we examined in the previous chapter. On the vertical axis is the percentage of GDP spent on healthcare and on the horizontal axis is average income.



Notice from the chart:

- Vietnam and Nigeria have roughly the same average incomes, but Vietnam spends over 7% of its GDP on healthcare while Nigeria spends less than 5% of its GDP on healthcare.
- Poland spends around 9% of its GDP on healthcare while Equatorial Guinea spends less than 5%.
- Norway spends around 10% of its GDP on health while Kuwait spends less than 3%.

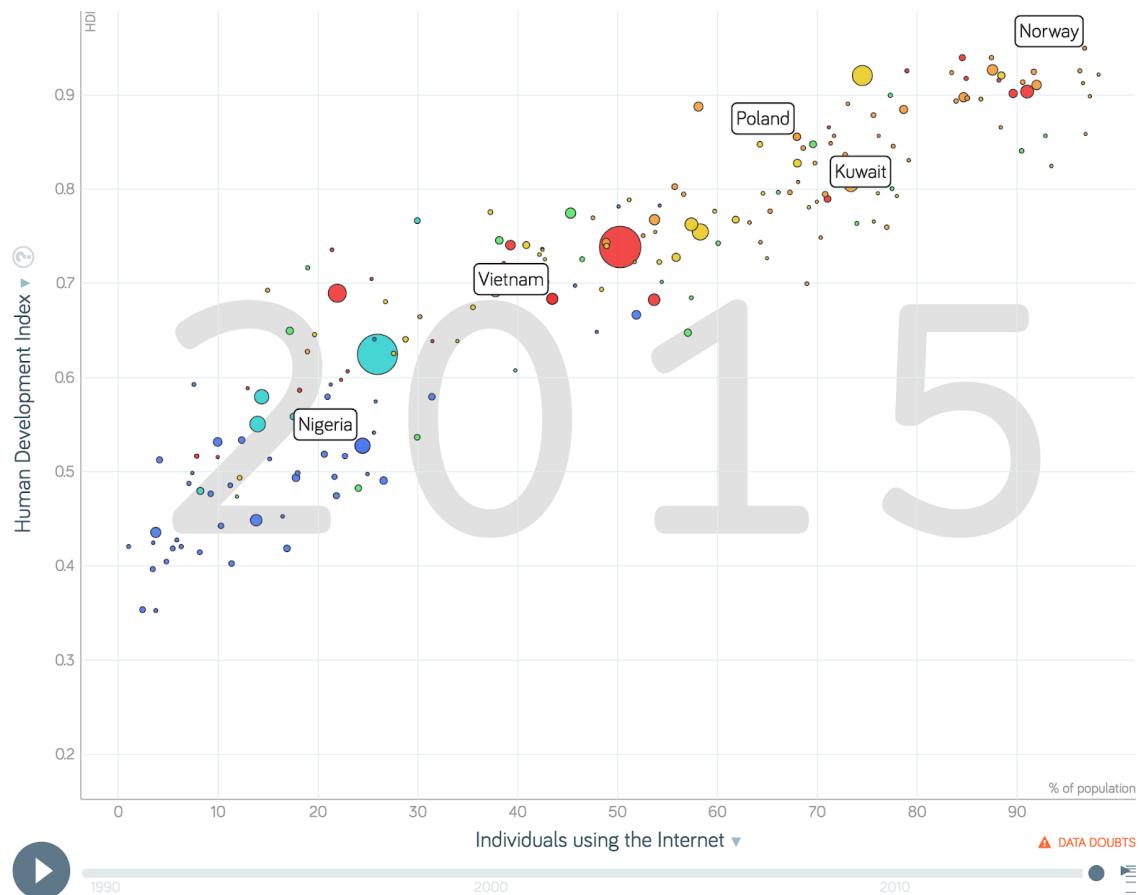
In each of the three cases, the country that spends a larger proportion of its GDP on health enjoys a higher HDI score (see previous chapter) than the country with roughly the same income that spends less on healthcare. Prioritizing the provision of healthcare, either through the free market or government intervention, helps countries of any income level enjoy higher life quality through longer lives and healthier population in general.

Technology

Health and education are not the only domestic factors that contribute to economic development. Access to appropriate technologies can also accelerate both the pace of economic growth and development for a country.

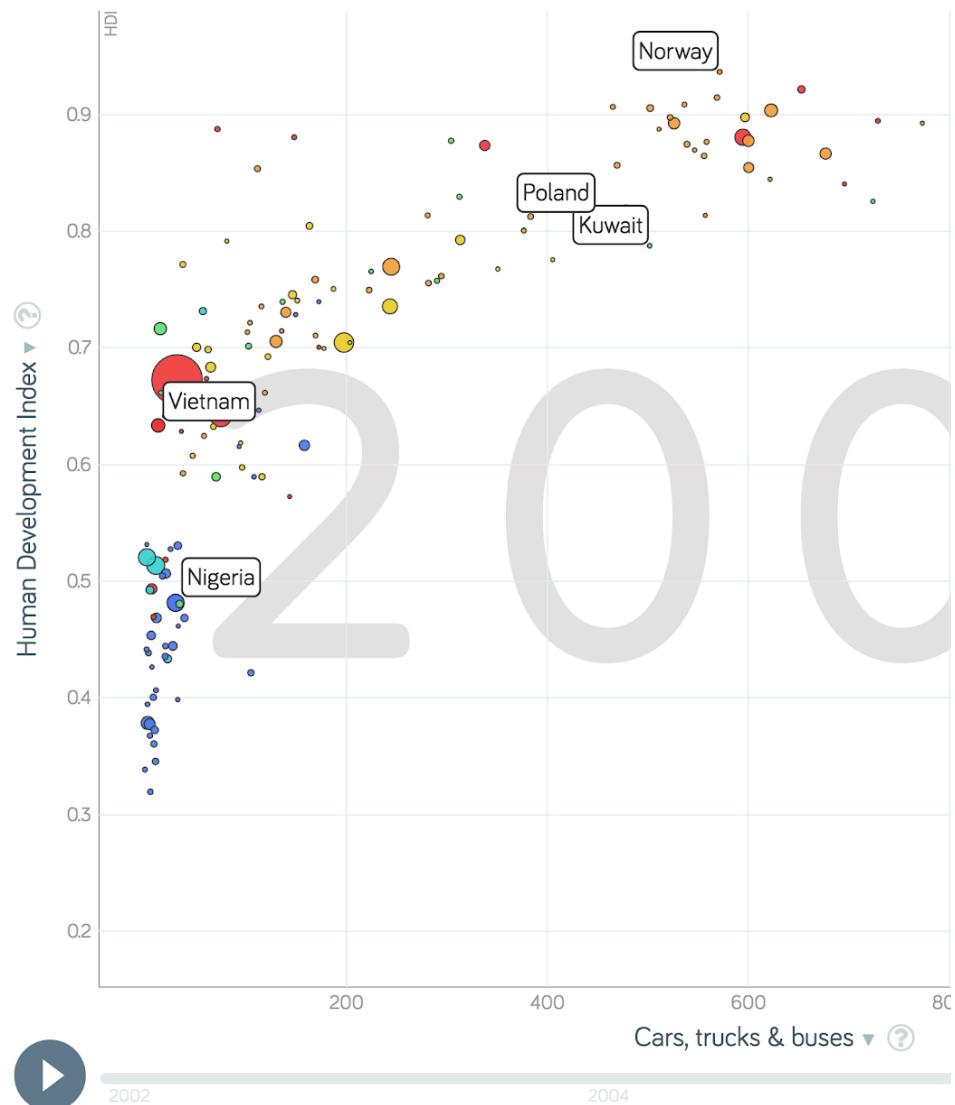
Communications technologies like broadband internet and cellular phones provide access to markets, information about prices to buyers and sellers, and improved access to the information necessary for producers to allocate their resources in an efficient manner in both local and global markets.

The chart below shows the number of people using the internet per 100 population (on the horizontal axis) against HDI score (on the vertical axis).



Notice from the chart that there is a strong correlation between internet usage and HDI score. Countries in which more people use the internet have a higher HDI score (Poland, Kuwait, Norway), while lower internet usage corresponds with lower HDI scores (Vietnam and Nigeria). The information and access to markets provided by the internet improve the efficiency of markets and help match producers with consumers (and vice versa). More efficient markets mean the goods and services necessary for economic growth and development are more likely to be provided.

In addition to communication technologies like the internet and cell phones, transportation technology can also help meet the growth and development needs of a country. The chart below shows the number of cars, trucks, and buses per 1,000 population (horizontal axis) plotted against HDI score (vertical axis).



Notice once again that there is a correlation between the quantity of transportation technologies and the level of economic development as measured by the HDI. In countries with very few cars, trucks, and buses HDI scores are relatively low, while more of these transportation goods correspond to higher HDI scores.

Transportation and communication technologies increase the efficiency of markets by making it easier for producers to access their consumers, be it through the provision of services online or through transporting physical goods via truck. Cars and buses allow workers to get to jobs wherever they may be located, thus contributing to employment opportunities and boosting incomes.

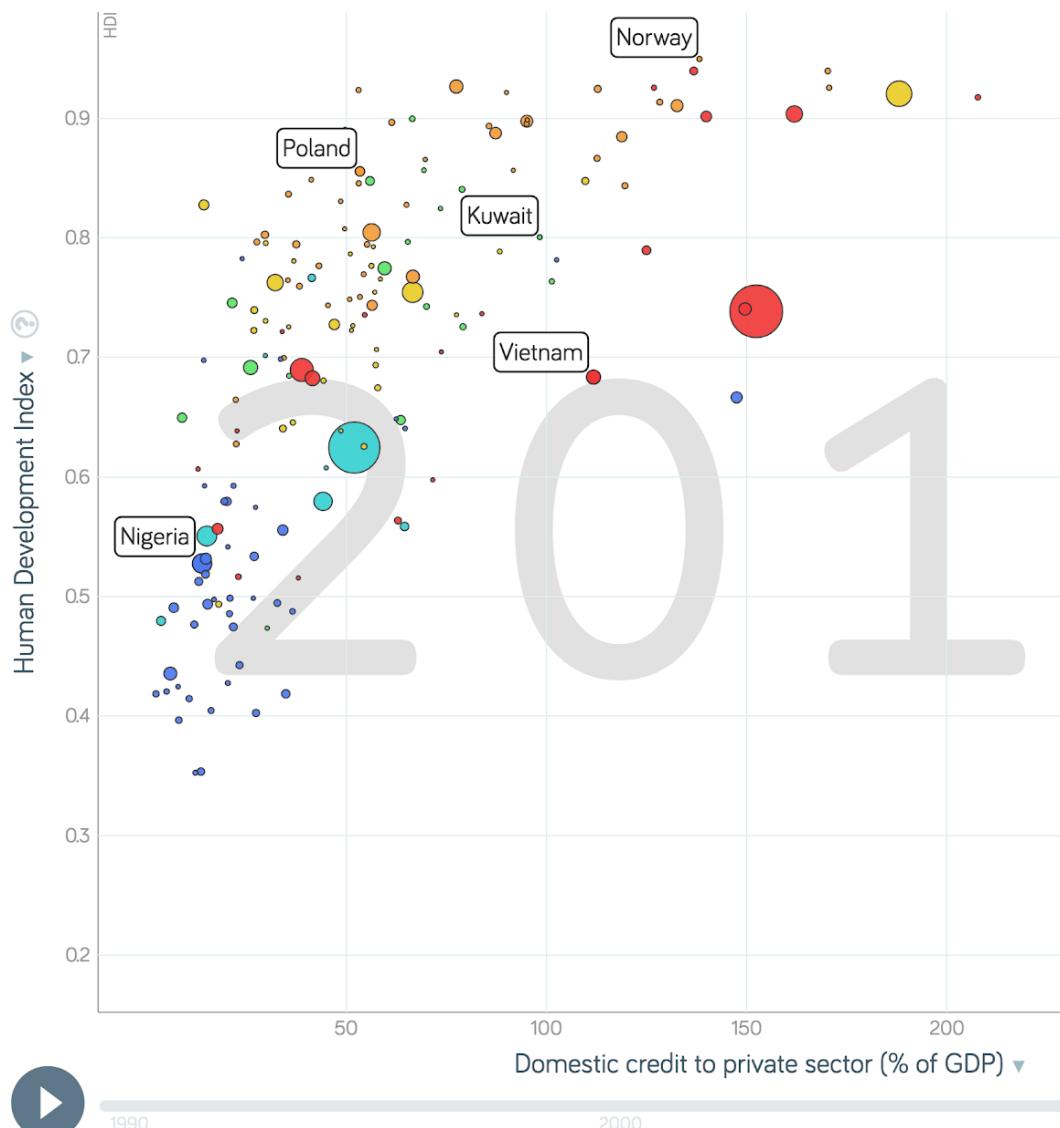
Credit and micro-credit

Modern economies depend on the efficiency of not only product markets, but resources markets as well. Resource markets include labor markets, the market for land, and the

market for capital. **Capital markets**, or **financial markets**, are where businesses that require capital are matched with institutions that make capital available.

Recall that **capital** is defined as the manufactured goods that are used in the production of other goods and services; it is the technology, the machines, the tools and robots that are used to produce other goods and services. In order to acquire capital, entrepreneurs often require **money** in order to invest in the tools and technology they need to start a business. The money used to invest in physical capital is referred to **financial capital**.

The ability of entrepreneurs to access financial capital is strongly correlated with a country's level of economic growth and development. One indicator of access to financial capital is the amount of private credit (loans made to private individuals and firms) as a percentage of a country's GDP. The chart below shows the amount of domestic credit to the private sector (horizontal axis) against HDI score (vertical axis).



Notice from the chart that once again there is a strong correlation between the two variables. Countries with more domestic credit (in other words, greater access to financial capital by domestic firms) enjoy higher HDI scores.

In most of Africa (blue dots), domestic credit to the private sector makes up less than 50% of GDP. This means that business owners in Africa have very little access to loans when they want to start or expand their businesses. The underprovision of credit is evidence of inefficient or nonexistent financial institutions. Banks and other lending institutions either do not exist or they do not extend much credit to the private sector.

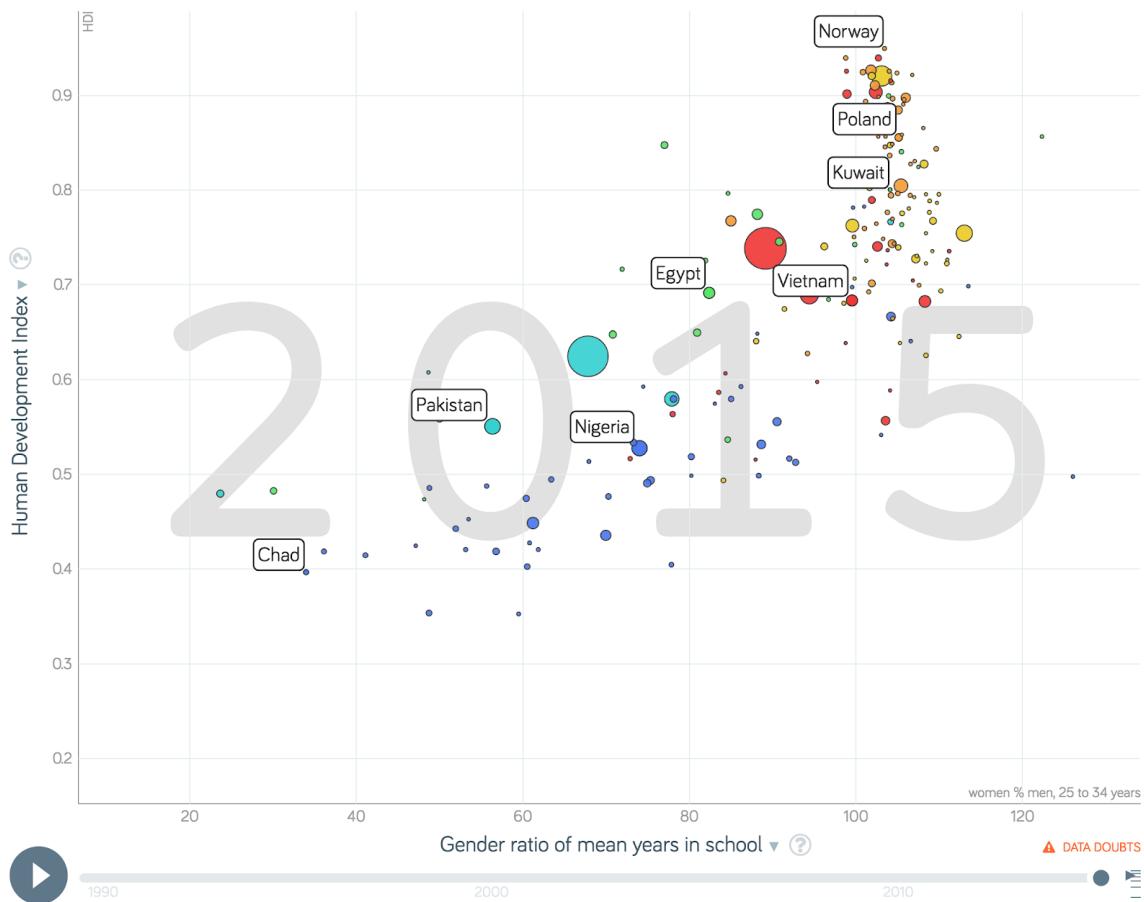
Greater access to financial capital can be achieved through the introduction of traditional banks to serve private business owners in LEDCs, or through the expansion of **microcredit** institutions, such as the Grameen Bank, which extends very small loans to poor borrowers who may lack collateral, formal employment, or a credit history. Such borrowers would be excluded from the traditional banking system, which is typically very conservative about who it makes loans to out of fear of not being repaid.

Microcredit institutions provide loans to individuals that otherwise would be unable to access the financial capital they need to start or expand a business. This expansion of credit has the potential to accelerate economic growth and development in LEDCs through increasing the amount of physical capital, expanding employment opportunities and the provision of welfare-enhancing goods and services to a country's population.

The empowerment of women

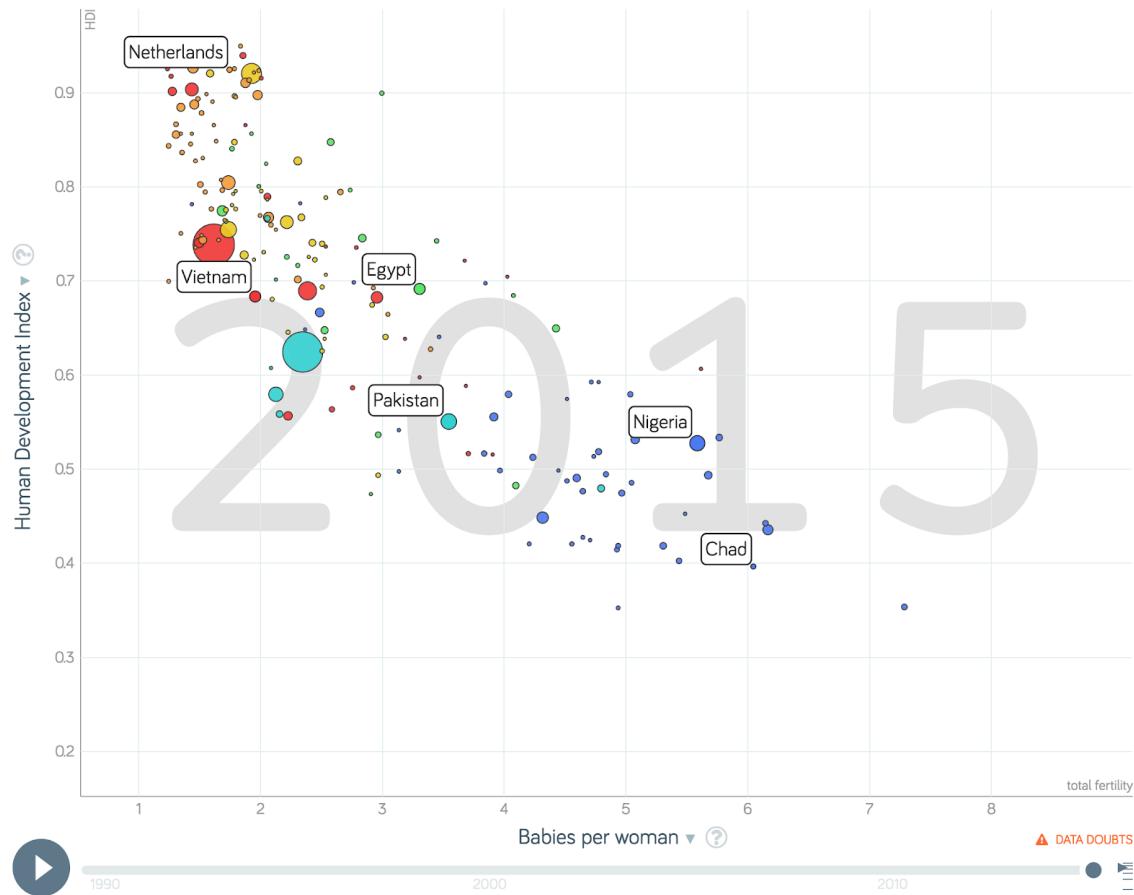
Another domestic factor affecting economic development is the role of women in society. Generally, the greater the role women play in contributing to economic activity in the country, the higher average incomes and the level of economic development.

The chart below shows the ratio of mean years to school of female to male. In other words, how many females per 100 males have received the average number of years of schooling in the country.



Observe from the chart that in some countries, such as Chad, fewer than 40 women per 100 men have achieved the average amount of schooling. Countries in which fewer women receive an education than men (to the left of 100 on the horizontal axis) generally have less gender equality than those around 100. Where women receive less education, they are less likely to contribute to the country's economy in a way that provides income and development opportunities.

Another indicator of gender equality is the total fertility rate, which measures the number of babies the typical woman will give birth to. The chart below looks at total fertility against HDI score.

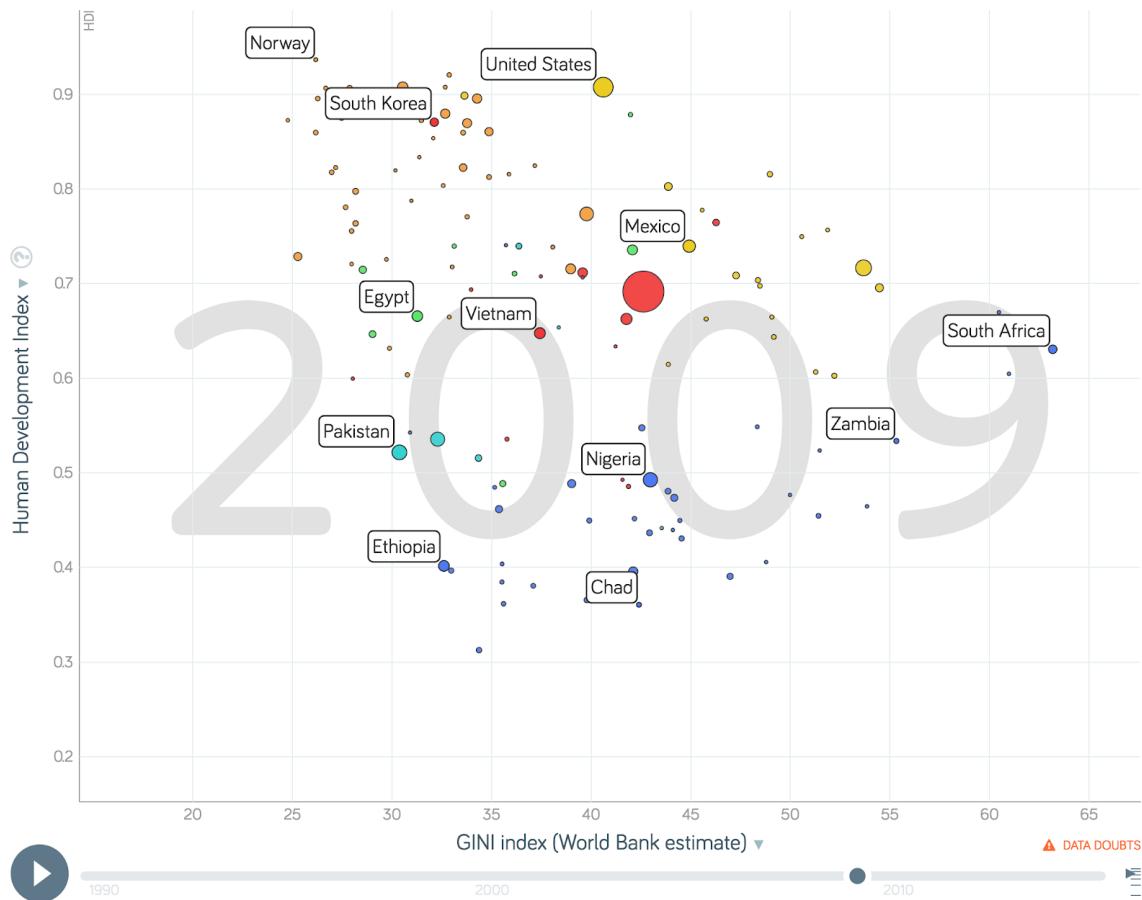


Notice the strong linearity in the chart. LEDCs in which women give birth to more than 4 children on average (including Nigeria and Chad) are have lower HDI scores than countries with lower fertility rates (like Vietnam and the Netherlands). When women are expected to give birth to more children, they are less likely to be participating in other activities such as receiving an education and earning income in the formal sector, both of which would contribute to economic development and growth.

Income distribution

For our final domestic factor affecting economic development we will consider income distribution. Recall from our study of macroeconomics that the level of income inequality in a country can be measured using the **Gini index**. The higher the Gini score, the less equally income is distributed; the lower the score, the more equal a country's income distribution.

Let's examine the relationship between Gini scores and HDI.



The first thing to observe about this chart is that there is not a general “trend line” that can be drawn showing the relationship between inequality and economic development. There are highly developed countries that are relatively equal (Norway and South Korea), but there are highly equal countries that are relatively undeveloped (Pakistan and Ethiopia). On the other hand, there are highly unequal countries that are relatively developed (The USA and Mexico) while there are also unequal countries that are relatively undeveloped (Zambia, Nigeria, South Africa).

Despite these observations we can make some general assumptions about the effect of inequality on economic development:

- A poor country that is highly unequal is likely to have low levels of economic development due to the lack of access to health, education, and other welfare-improving goods and services for a large proportion of the population.
- Rich countries that are relatively equal are likely to have higher HDI scores than rich countries that are highly unequal. Norway scores better than the US on the HDI rankings due to the fact that more of its people have access to quality education and health than in the US, where greater inequality means there is more relative poverty.

High degrees of inequality can pose an obstacle to economic development because a large percentage of the population is excluded from the gains from economic growth. Steps to

reduce inequality (progressive taxes, transfer payments, the provision of public and merit goods) can assure that more people in society enjoy the gains of economic growth and development.

On the other hand, a government that implements too aggressive a program of income redistribution can inadvertently disincentivize entrepreneurship, hard work, and business investment, leading to **capital flight** and **brain drain**, in which domestic firms relocate their operations outside the country and skilled workers seek employment abroad, leaving the country with less physical and human capital, both of which are necessary ingredients for economic growth and development.

4.4 The Role of International Trade in Economic Development

Trade problems facing many LEDCs

- With reference to specific examples, explain how the following factors are barriers to development for economically less developed countries.
 - a. Over-specialization on a narrow range of products
 - b. Price volatility of primary products
 - c. Inability to access international markets

HL only objective:

- With reference to specific examples, evaluate long-term changes in the terms of trade as a barrier to development for economically less developed countries.

Note: The charts in this chapter were created using MIT's Observatory of Economic Complexity and the St. Louis Federal Reserve Bank's FRED database. Both resources can be used to create highly customized interactive charts presenting economic data.

In a globalized world economic development requires more than just domestic measures such as improvements to education and health, improvements to technology, access to credit, and the empowerment of women. A country's role in the international markets and what it has to offer its trading partners in terms of goods for export will also have a significant impact on its level of economic development.

In the international trade section of this course we learned about the benefits countries enjoy from specializing in the production of the goods in which they have a **comparative advantage**. While specialization allows countries to achieve greater efficiency and to benefit from one another's lower opportunity costs in various goods' production, taken to its extreme **over-specialization** can have dire consequences for a country's economic growth and development.

Many less economically developed countries tend to over-specialize in a narrow range of products, often-times primary commodities such as energy resources, minerals, and agricultural goods. Consider the graph below, which shows the composition of Nigeria's exports to the rest of the world.

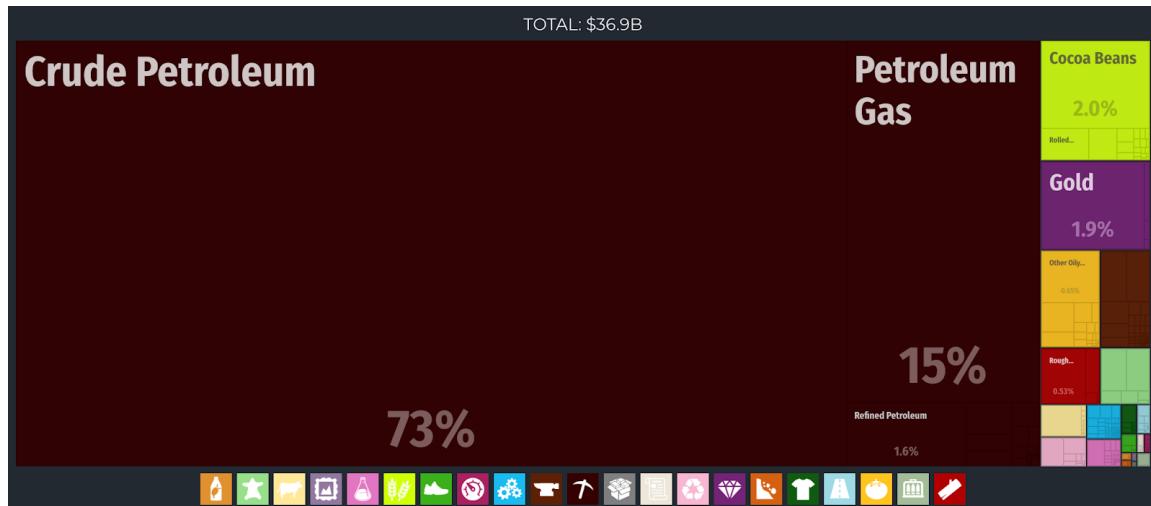


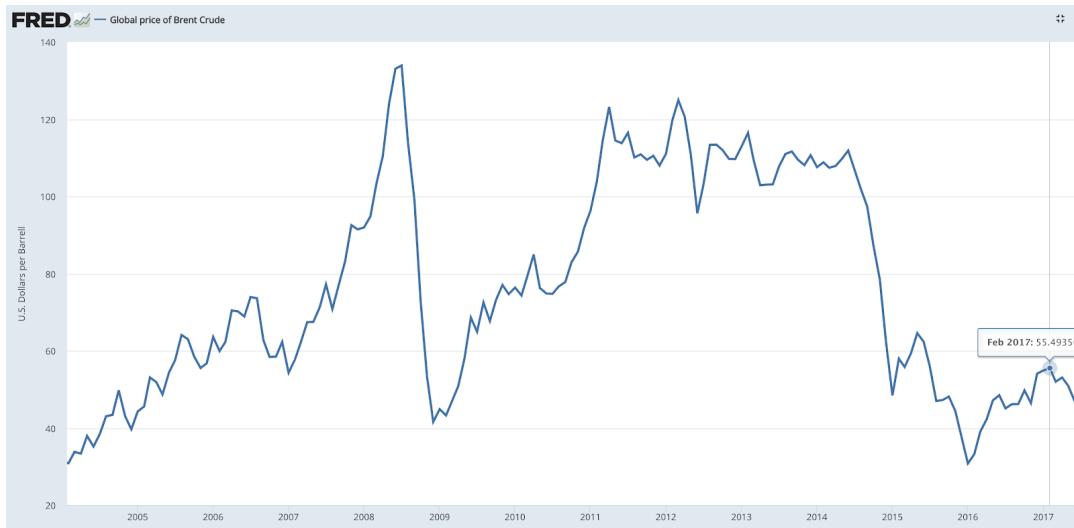
Figure 1: Nigeria exports⁶

Observe from the graph that 88% of Nigeria's exports are of energy resources (oil and gas). While specialization in petroleum allows Nigeria to achieve a high level of efficiency in the production of this valuable resource, it also leaves the country highly vulnerable to fluctuations in global petroleum prices.

When oil prices are high, Nigeria can expect to enjoy booming export revenues, strong aggregate demand, higher levels of employment and income, and a strong currency. However, when global oil prices fall, Nigeria's aggregate demand, employment, price level, aggregate output, and currency exchange rate will all decline. The current account balance will move towards deficit and Nigeria's foreign income will fall, making it difficult to afford the imports of manufactured goods and technology that it depends on for economic development.

Consider the chart below, which shows the world oil price between 2005 and 2017.

⁶ Source: The Observatory of Economic Complexity - <https://atlas.media.mit.edu/en/>

Figure 2: Global oil price, 2005-2017⁷

As the price of its only major export fluctuated, Nigeria's economy would have experienced macroeconomic shocks as its export earnings rose and fell. For example, consider the chart below, which shows Nigeria's terms of trade (HL only concept) over the same period.

Figure 3: Nigeria's terms of trade, 2005-2018⁸

Terms of trade represents the amount of imports a country can buy per unit of exports it sells. Notice that Nigeria's terms of trade rise and fall almost lockstep with oil prices. When its terms of trade improve, Nigeria finds imported technology and consumer goods relatively cheap, allowing for improvements in economic development as the cost of welfare-improving products become more affordable.

⁷ Source: FRED Economic data - <https://fred.stlouisfed.org/series/POILBREUSDM>

⁸ Source: FRED Economic data - <https://fred.stlouisfed.org/series/NGATTIX>

However, when oil prices fall, Nigeria's terms of trade deteriorate, as they did between 2014 and 2016 when terms of trade fell from 114 to 76, a 33% decline. In other words, between 2014 and 2016 the amount of imports Nigeria could buy per unit of its exports fell by 33%.

Let's consider another country that is overly-dependent on a narrow range of exports. The graph below shows the composition of Malawi's exports.



Figure 4: Malawi's exports⁹

Malawi specializes almost entirely in agricultural goods, which like oil tend to have highly volatile prices on global markets due to their highly inelastic supply and demand. Consumers are not highly responsive to price changes, so changes in supply year to year tend to result in sharp spikes or dips in global price. In the short-run, supply of agriculture goods is highly inelastic, resulting in price volatility as demand rises and falls.

The inelastic supply and demand for agricultural goods leads to price volatility, resulting in instability in countries that specialize in such goods' terms of trade. The chart below shows Malawi's terms of trade (HL only concept) between 2005 and 2018.

⁹ Source: Observatory of Economic Complexity - <https://atlas.media.mit.edu/en/>

Figure 5: Malawi's terms of trade, 2005-2018¹⁰

Malawi's terms of trade fluctuated from a low of 75 to a high of 100, rising by 33% in just a year and then falling by 25% over the following years. This volatility in the relative price of exports to imports reflects the price instability of its agricultural exports. And like Nigeria, for a country that depends on the import of nearly all manufactured goods, technology, medical instruments, drugs, and other welfare-improving goods, a narrow range of exports and the resulting volatility in terms of trade poses an obstacle to economic development.

For comparison, let's look at the composition of exports for a more economically developed country, the United Kingdom. The graph below shows the UK's exports by category.

Figure 6: United Kingdom exports¹¹

¹⁰ Source: FRED Economic data - <https://fred.stlouisfed.org/series/MWITTTIX>

¹¹ Source: Observatory of Economic Complexity - <https://atlas.media.mit.edu/en/>

Obviously, the UK produces a much more diverse range of goods for export to the rest of the world. Let's next see how the UK's terms of trade changed over the same period of time we observed in Nigeria and Malawi.



Figure 7: United Kingdom terms of trade¹²

While its terms of trade certainly fluctuated, you'll notice that the range over which it changed is much narrower than that of Nigeria and Malawi, falling from 101 in 2005 to 95 in 2008 (a decrease of only 6%) rising to 102 in 2016 (an approximately 7% increase). For comparison, Nigeria's and Malawi's terms of trade fluctuated by 33% over the same period.

Diversifying its output across a wide range of both primary and secondary (manufactured) goods shelters a country from the volatility in both export earnings and the ability to buy imports that result from over-specialization in a narrow range of exports.

Trade strategies for economic growth and economic development

- With reference to specific examples, evaluate each of the following as a means of achieving economic growth and economic development.
 - a. Import substitution
 - b. Export promotion
 - c. Trade liberalization
 - d. The role of the WTO
 - e. Bilateral and regional preferential trade agreements
 - f. Diversification

¹² Source: FRED Economic data - <https://fred.stlouisfed.org/series/TVTTUKA>

Several strategies exist in the area of international trade for countries hoping to achieve economic growth and economic development. In this section we will examine several of these strategies; we will define them, give examples of countries that have pursued them, and evaluate the extent to which they can be effective at promoting growth and development.

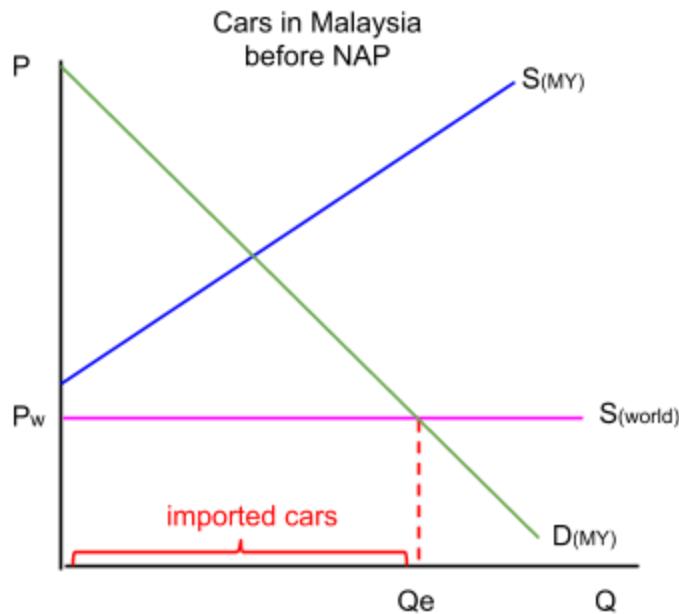
Import substitution

When a country uses **protectionist trade policies**, such as import tariffs, quotas, and subsidies, in order to encourage households to substitute domestically produced goods for imported goods, it is pursuing an **import substitution** policy of economic growth.

For example, consider Malaysia, which in the 1970s imported 100% of the automobiles on its roads. In the early 1980s the government decided that Malaysia should have its own automotive manufacturing industry, so it implemented the National Automotive Policy (NAP), which imposed tariffs of over 100% on most imported cars, while providing heavy subsidies to a domestic car manufacturing startup, Proton.

By the late 80's most of the imported cars on the road had been replaced by Protons, which could be bought far more cheaply by Malaysian consumers than the heavily taxed imports.

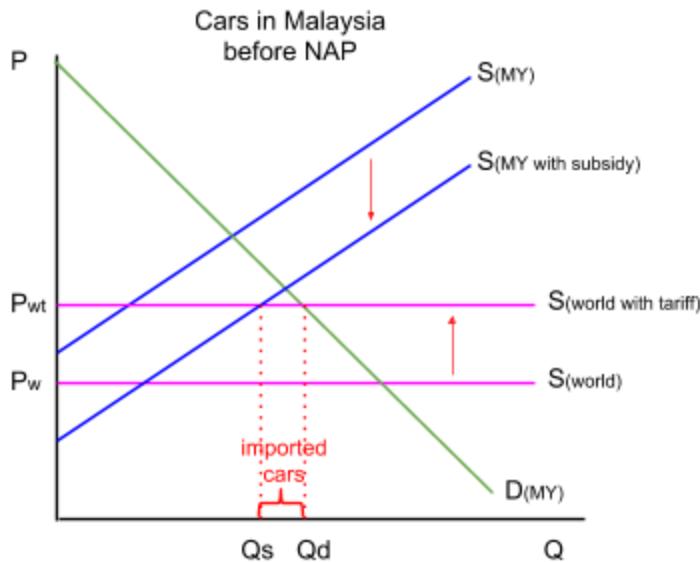
The effect of an import substitution strategy for economic growth can be illustrated using a demand and supply diagram. Consider the market for automobiles before Malaysia's NAP.



The world price is below the price at which Malaysian producers would even be willing or able to enter the car market (in other words, P_w is below the price-intercept of supply).

Following the implementation of the NAP, a tariff is imposed on imported cars at the same time that a subsidy is provided to Proton, the government-funded Malaysian auto

manufacturer. The effects of this import substitution policy can be illustrated on our graph.



Observe from the graph:

- The tariff has raised the price of imported cars from P_w to P_{wt} .
- The subsidy has lowered the costs (thus increased the supply) of Malaysian cars from $S(MY)$ to $S(MY \text{ with subsidy})$.
- At the new world price of P_{wt} the domestic firm is willing to produce at a quantity of Q_s , while the quantity demanded is Q_d .
- The quantity of imported cars has fallen from ALL cars before the NAP to only $Q_d - Q_s$ cars after the NAP.

Malaysia's National Automotive Policy is just one example of an import substitution approach to economic growth. The benefits to Malaysia's economy are clear: Where there was previously no automotive sector, there are now thousands of jobs in dozens of factories around the country producing cars for both the domestic and international market.

On the other hand, there are several disadvantages to an import substitution approach to economic growth:

- Malaysian consumers face much higher car prices than they would without the 100+% tariff on imported cars. Japanese, Korean, European, and even American cars all would have been cheaper and thus more attractive to Malaysian consumers had the NAP never been adopted.
- Foreign car producers suffer as their exports to Malaysia fall, reducing output and employment in more efficient foreign automobile markets.
- Subsidies for Malaysia's car producer come at a large cost to Malaysian taxpayers, whose hard earned incomes are used to subsidize a relatively inefficient domestic producer.
- Resources are misallocated both in Malaysia and in countries where cars can be produced more efficiently. There is a loss of total surplus (welfare) and allocative

- inefficiency as a result of the protectionist policies.
- Foreign governments may retaliate against Malaysia's protectionist measures, harming producers of other goods that Malaysia may actually have a comparative advantage in the production of.

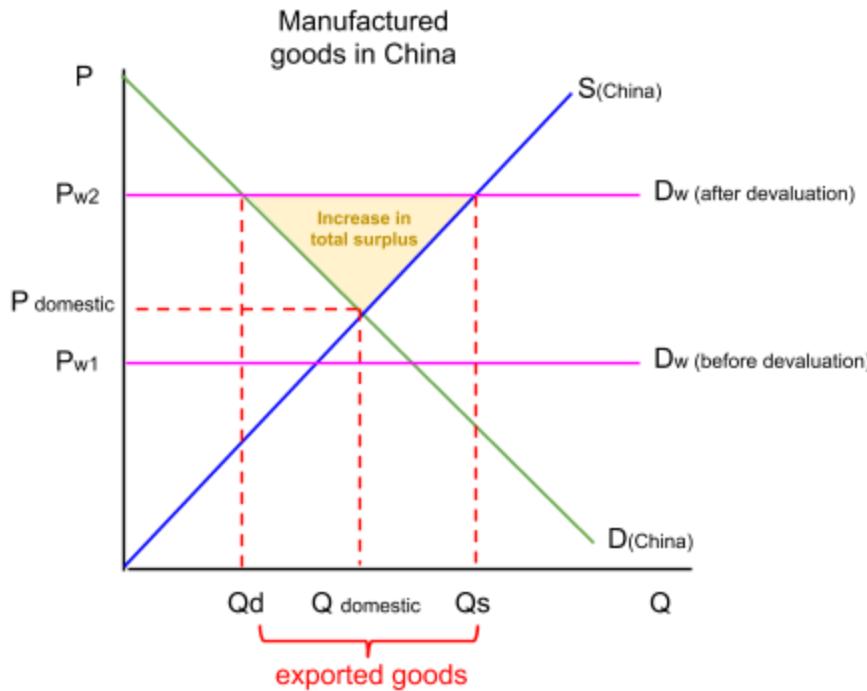
Import substitution could provide LEDCs with a useful means of protecting and growing **infant industries**, or those that have yet to achieve the **economies of scale** necessary to compete with larger foreign producers. But they should be used only temporarily. Once the domestic industries are large and efficient enough to compete with foreign firms, protections should be rolled back, allowing domestic and foreign firms to compete on a level playing field.

Export promotion

Another approach to achieving economic growth and development is for a country to adopt an **export promotion** strategy. Export promotion entails subsidizing domestic firms and manipulating the currency in ways that allow them to achieve lower costs than foreign producers, exporting surplus output onto the global market, and thus boosting domestic output and employment.

An example of an export promotion strategy for economic growth is the path China pursued during much of the first decade of the 21st century. During the 2000s, China employed an aggressive policy of **fixed exchange rates** that required competitive **devaluation** of the Chinese yuan on forex markets. By fixing its exchange rate at below market values, particularly against the US dollar, China was able to achieve and maintain persistent current account surpluses with most of its trading partners.

In the graph below we can see the effect on China's market for manufactured goods before and after the adoption of a fixed exchange rate for the Chinese yuan.



Observe from the graph:

- Before the yuan's devaluation the world demand for Chinese manufactured goods was relatively low. The domestic price (P_{domestic}) was higher than the world price (P_{w1}), so Chinese producers produced just enough goods for Chinese consumers (Q_{domestic}). There were no exports from China to the rest of the world.
- After the adoption of a fixed exchange rate that resulted in the devaluation of the yuan, global demand for China's manufactured goods increased to D_w (after devaluation). The cheaper yuan made Chinese goods far more attractive to foreign consumers.
- Increased demand from the rest of the world drove the price of manufactured goods in China up, increasing the quantity supplied to Q_s and reducing the quantity demanded to Q_d .
- China now exports manufactured goods to the rest of the world and there is an increase in total surplus in China represented by the yellow triangle.

China's export promotion strategy for economic growth has helped lift hundreds of millions of the country's people out of poverty and resulted in the largest migration of people in modern history as the country's population has shifted from rural, agricultural areas to the urban, manufacturing hubs along the country's coast. Today China enjoys the world's second highest GDP and the most exports of any country in the world.

As with other growth and development strategies, export promotion comes with its risks as well. The manipulation of exchange rates and the use of subsidies to support domestic firms runs afoul of WTO rules governing trade between member countries (which China has been since 2001). Trading partners may view export promotion as an "unfair" trade strategy and

retaliate with protectionist tariffs of their own.

Furthermore, subsidies and forex market intervention lead to a misallocation of world resources as the export-promoting country will produce more of certain goods than it would under free trade, when resources are allocated purely by who has the lower opportunity cost in different goods' production.

4.5 The Role of Foreign Direct Investment (FDI)

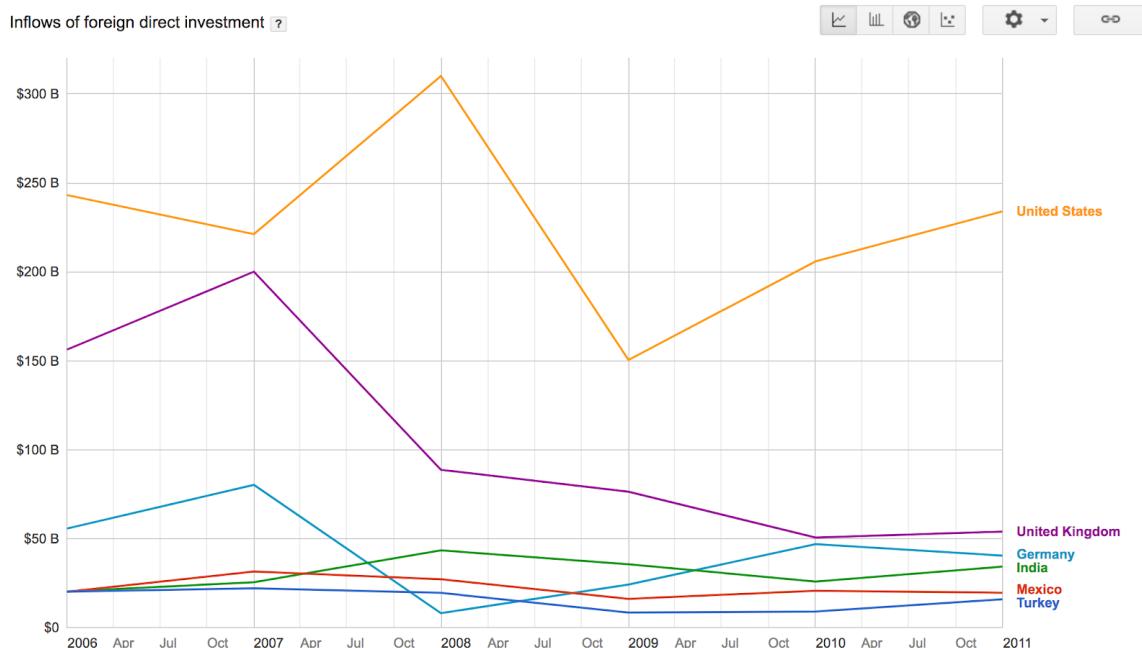
The meaning of FDI and multinational corporations (MNCs)

- Describe the nature of foreign direct investment (FDI) and multinational corporations (MNCs).
- Explain the reasons why MNCs expand into economically less developed countries.
- Explain the characteristics of economically less developed countries that attract FDI, including low cost factor inputs, a regulatory framework that favours profit repatriation and favourable tax rules

According to the United Nations Conference on Trade and Development (UNCTAD) **foreign direct investment (FDI)** “refers to an investment made to acquire lasting interest in enterprises operating outside of the economy of the investor.¹³” FDI occurs when an investor, usually a **multinational corporation (MNC)** acquires at least a 10% stake in a business outside of its home country.

MNCs engage in FDI in both more and less economically developed countries. By far vast majority of FDI flows from rich countries into other rich countries.

Examine the chart below, which shows FDI inflows in six countries between 2006 and 2011.

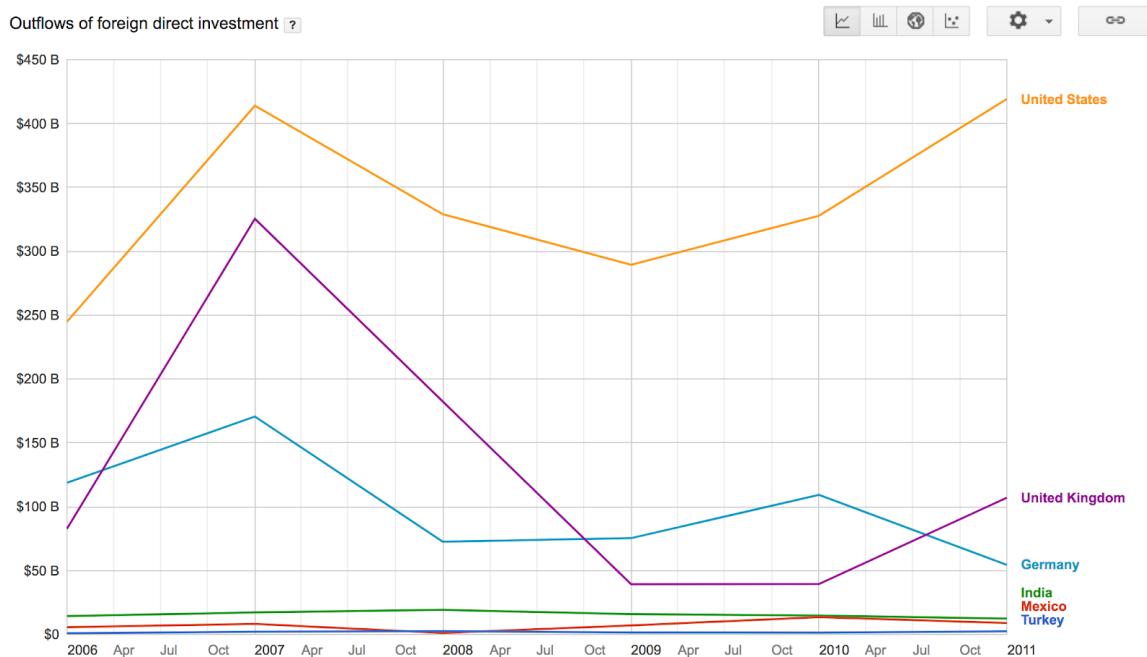


Notice from the chart that the three MEDCs (the US, the UK, and Germany) received more total investment resulting in foreign ownership of domestic enterprises than the three

¹³ Source: [http://unctad.org/en/Pages/DIAE/Foreign-Direct-Investment-\(FDI\).aspx](http://unctad.org/en/Pages/DIAE/Foreign-Direct-Investment-(FDI).aspx)

LEDCs (India, Mexico, and Turkey). This simply reflects the fact that larger, richer, economies have more opportunities for investment by foreign firms than smaller, poorer economies.

While MEDCs are the largest recipients of FDI, they are also its largest sources. Given that the largest MNCs are based in more economically developed countries, this should not come as a surprise. The chart below shows the outflow of FDI in the same six countries.



While American firms invested over \$400 billion in foreign enterprises in 2011, Turkey's firms invested only \$2.4 billion, Mexico's \$8.9 billion, and India's \$12.4 billion. German and British firms invested \$54.4 billion and \$107 billion abroad, respectively.

In examining the role of FDI in economic development, we must consider the reasons why MNCs expand into economically less developed countries. Investing in businesses in relatively poor, underdeveloped countries can provide MNCs with several benefits, including:

- **Access to markets** - While the people living in LEDCs are generally poorer than those in MEDCs, they do make up the majority of the world's population, and their incomes are growing faster than incomes in the rich world. The emerging middle class in many LEDCs presents a desirable consumer base that MNCs producing consumer goods are eager to tap into. Having a presence in the countries where the bulk of the world's population lives allows companies to sell their goods to an ever-growing base of middle class households.
- **Low cost labor** - While most of the world's FDI flows from rich countries to other rich countries, the low wages in many LEDCs makes them an attractive place to locate factories, farms, and other production facilities that require large amounts of

- labor.
- **Raw materials** - Many of the world's most valuable natural resources are found in great abundance in less economically developed countries. From scarce minerals like "rare-earth metals", which are essential for the production of modern consumer electronics like smart phones, to abundant timber supplies, to land for grazing livestock and farming soybeans, LEDCs often possess natural resources that are relatively scarce in MNCs' home countries, making them attractive destinations of foreign direct investment.

Advantages and disadvantages of FDI for economically less developed countries

- Evaluate the impact of foreign direct investment (FDI) for economically less developed countries.

Multinational corporations invest in less economically developed countries because doing so benefits the corporations. Whether or not such investment also benefits the countries receiving it depends on several factors.

Let's examine some of the characteristics of LEDCs that may attract FDI in the first place:

- **A lax regulatory environment** - many LEDCs lack a regulatory framework that effectively oversees means of manufacturing of goods. Lax environmental and labor protections may make LEDCs a cheap and desirable place to produce goods that require low-cost labor and create pollution in their production. If workers are unfairly exploited and the environment is degraded as a result of the activities of foreign-owned businesses, then economic development may be sacrificed in the pursuit of profits.
- **Favorable tax rules** - There are over 150 countries classified as LEDCs in the world today. Most of these would love to attract foreign investment because of the jobs and income it creates. In an attempt to make their countries more attractive to large multinational corporations, governments may get caught up in a "race to the bottom" by offering deep tax breaks or even subsidies to foreign firms. As a result, an inflow of FDI may not lead to more tax revenues for the poor country's government, limiting its ability to provide public goods like infrastructure, education and healthcare. The foreign investor enjoys the benefits of low taxes but the country's people miss out on the benefits of having a large corporation contributing to the country's tax base.
- **Ability to repatriate profits** - LEDCs that manage to attract foreign firms, then tax their earnings to provide public goods, may find their tax revenues still falling short due to loopholes in the international tax structures that allow multinational firms to report earnings in whichever country offers the lowest taxes, regardless of where the goods are actually produced. A practice known as "**transfer pricing**" demonstrates how this works. Transfer pricing allows MNCs to pay little or no tax in the poor countries in which they operate in the following way:
 - Assume a large mine in a sub-Saharan African country is majority-owned by a mining company based in Switzerland (a European country with some of the lowest taxes in the world).

- When the African mine goes to sell its output, it will sell to its parent company in Switzerland at a price far below the commodity's price on world markets.
- The mine then reports its earnings to the government of the country in which it is located, and it turns out the mine only broke even or earned a loss. As a result, little or no tax is paid to the African country's government.
- The Swiss parent company resells the minerals on the global market at market prices, earning the firm a large profit.
- Taxes on the MNC's profits are paid to the Swiss government, where the profits are actually earned. As a result, Switzerland, a very rich country, enjoys higher tax revenues while the sub-Saharan African country remains trapped in poverty.

As we can see, FDI in less economically developed countries comes with its advantages and disadvantages. The table below summarizes these.

Potential benefits of FDI for LEDCs	Potential costs of FDI for LEDCs
More employment	Workers are exploited and the environment degraded, or jobs go to foreign workers brought into country
Higher incomes	Income is transferred abroad or earned by foreign workers
More tax revenues	Profits are repatriated and taxes paid abroad
New technology and infrastructure	New technology and infrastructure is limited and only benefit foreign firms

In conclusion, it should be pointed out that the vast majority of foreign direct investment is in sectors that produce consumer goods or primary commodities. Multinational corporations generally are not interested in investing in poor countries' health and education sectors, or in funding infrastructure that does not directly benefit the company itself. For these reasons, FDI is more likely to contribute to economic growth than to economic development.

4.6 The Roles of Foreign Aid and Multilateral Development Assistance

Classifications and types of aid

- Explain that aid is extended to economically less developed countries either by governments of donor countries, in which case it is called official development assistance (ODA), or by nongovernmental organizations (NGOs).
- Explain that humanitarian aid consists of food aid, medical aid and emergency relief aid.
- Explain that development aid consists of grants, concessional long-term loans, project aid that includes support for schools and hospitals, and programme aid that includes support for sectors such as the education sector and the financial sector.
- Explain that, for the most part, the priority of NGOs is to provide aid on a small scale to achieve development objectives.
- Explain that aid might also come in the form of tied aid.
- Examine the motivations of economically more developed countries giving aid.
- Compare and contrast the extent, nature and sources of ODA to two economically less developed countries.

Foreign aid is defined simply as, “money, food, or other resources given or lent by one country to another.” Notice that aid can be “given” (meaning there is no expectation of repayment) or “lent” (in which case, the recipient country is required to repay some of what is given to the donor country).

The vast majority of the over \$150 billion in aid given annually is in the form of **official development assistance (ODA)** which is aid extended by one government (the donor) or from multilateral development institutions such as the World Bank and the International Monetary Fund (IMF) to another government (the recipient).

Aid is also extended, but at a much smaller level, by **nongovernmental organizations (NGOs)**. Some well-known examples of NGOs providing aid to economically less developed countries, along with their primary focuses, include¹⁴:

- The Bill and Melinda Gates Foundation - global health
- World Vision - food aid and emergency assistance
- Oxfam - poverty alleviation and debt relief
- Doctors without Borders - healthcare services
- CARE International - poverty alleviation
- International Committee of the Red Cross - humanitarian relief

While these well-known organizations target specific areas of economic development, aid in

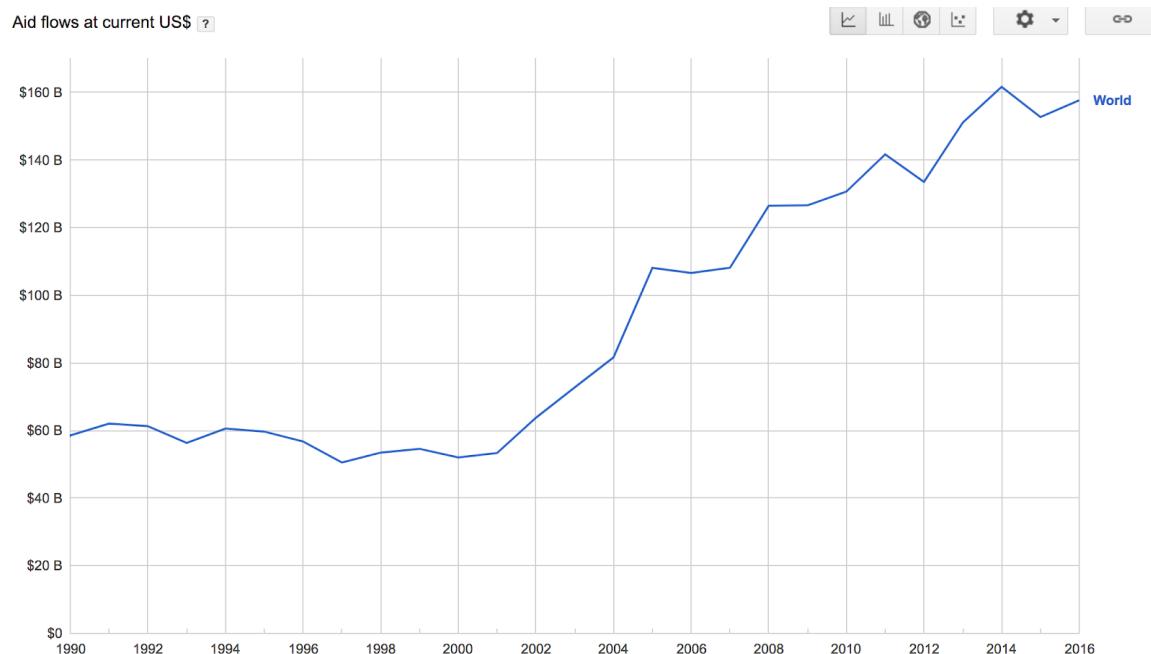
¹⁴ Source: “The World’s Most Powerful Development NGOs”

<https://foreignpolicy.com/2008/07/01/the-list-the-worlds-most-powerful-development-ngos/>

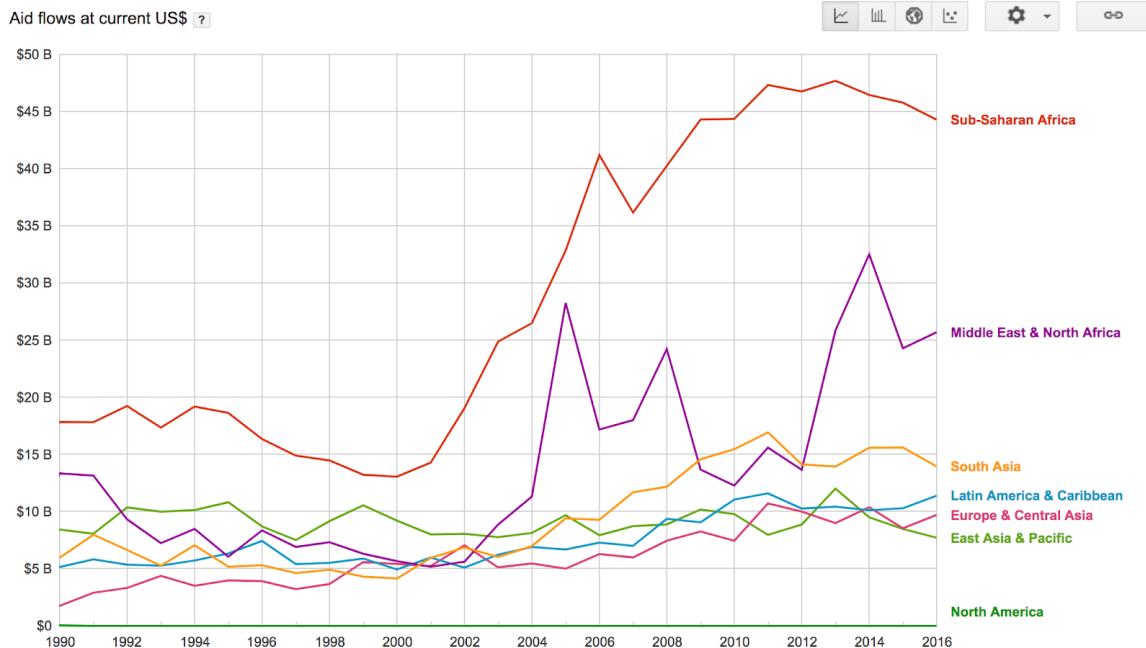
general falls into one of two categories, summarized in the table below.

Foreign Aid	
Humanitarian Aid	Development Aid
Examples: food aid, medical aid, disaster relief aid	Examples: grants, concessional loans, project aid, programme aid

Most official development assistance (government to government) is in the form of development aid. The chart below shows the dollar value of ODA extended worldwide from 1990 to 2016.



At last count over \$150 billion per year was being extended in foreign aid, mostly from rich countries to poor countries. When we look at aid flows by region, we can see where most of the money given for development assistance is flowing to.



Notice from the chart that the region receiving by far the largest proportion of foreign aid is the 54 countries of sub-Saharan Africa, where around 1 billion people live and where average incomes are lower than in any other region of the world.

Development aid comes in many forms. Below we'll define each:

- **Grants** are money gifted by a government to another government for the purpose of improving the country's infrastructure or other factor contributing to economic development. Grants are not required to be repaid by countries that receive them.
- **Concessional loans** are money lent by a government to another government, at least 25% of which does not have to be repaid. A loan is "concessional" in nature when the interest rate is below what the recipient country could receive from a private lender, which classifies such loans as "aid."
- **Project aid** is money given for specific investments in healthcare, education, or other infrastructure (hospitals, schools, ports, railways, bridges, etc...)
- **Programme aid** is money given to a country that is not designated for a specific project, including money to help balance a government's budget or to stabilize its currency's value on the forex markets.

Tied aid

According to the Organization for Economic Cooperation and Development (OECD), **tied aid** describes "official grants or loans that limit procurement to companies in the donor country or in a small group of countries.¹⁵" In other words, when a government makes a grant or a loan to another government, but requires the recipient to use the money to buy technology or hire services from companies in the donor country.

¹⁵ Source: "Untying aid: the right to choose"

<http://www.oecd.org/development/untingaidtherighttochoose.htm>

Tied aid is aid with “strings attached.” It effectively limits the ability of the recipients to “shop around” for the best value for the goods, services, or work it wishes to invest in with the granted or loaned funds. Tied aid acts as an indirect stimulus to the donor country’s economy, masked as development assistance for a poorer country. Since the money has to be spent in the donor country, it boosts domestic output and employment at the expense of the recipient having the freedom to choose how where and how to spend the money in a way that maximizes its impact on the country’s development.

According to the OECD, the proportion of aid that is “untied” (e.g. the recipient is able to spend it wherever it gets the most bang for the buck) has increased from 46% to 82%, vastly increasing its potential impact for good in the recipient countries.

Tied aid provides an obvious benefit for the donor country, but untied aid can stand to benefit both the recipients and the donors. What may appear as a “selfless act” can in fact benefit both sides of the aid transaction. Donor countries (mostly rich, more economically developed countries) are better off when the rest of the world experiences rising incomes, longer lives, and greater levels of education and health.

Motivations for donor countries

Improving development in poor countries creates more consumers for rich country firms, a better educated workforce willing and able to contribute to global production, and greater economic, political, and social stability, all of which contribute to favorable conditions for economic growth and increased revenues for firms and higher incomes for households in rich and poor countries alike.

Foreign aid case study #1 - Vietnam

In 2016 Vietnam received \$2.37 billion in official development assistance, making it the world’s third largest recipient nation after Afghanistan and India¹⁶. The graphic below breaks down the extent, nature, and sources of the ODA received by Vietnam in 2016¹⁷.

¹⁶ Source: “Aid at a glance charts” from the OECD

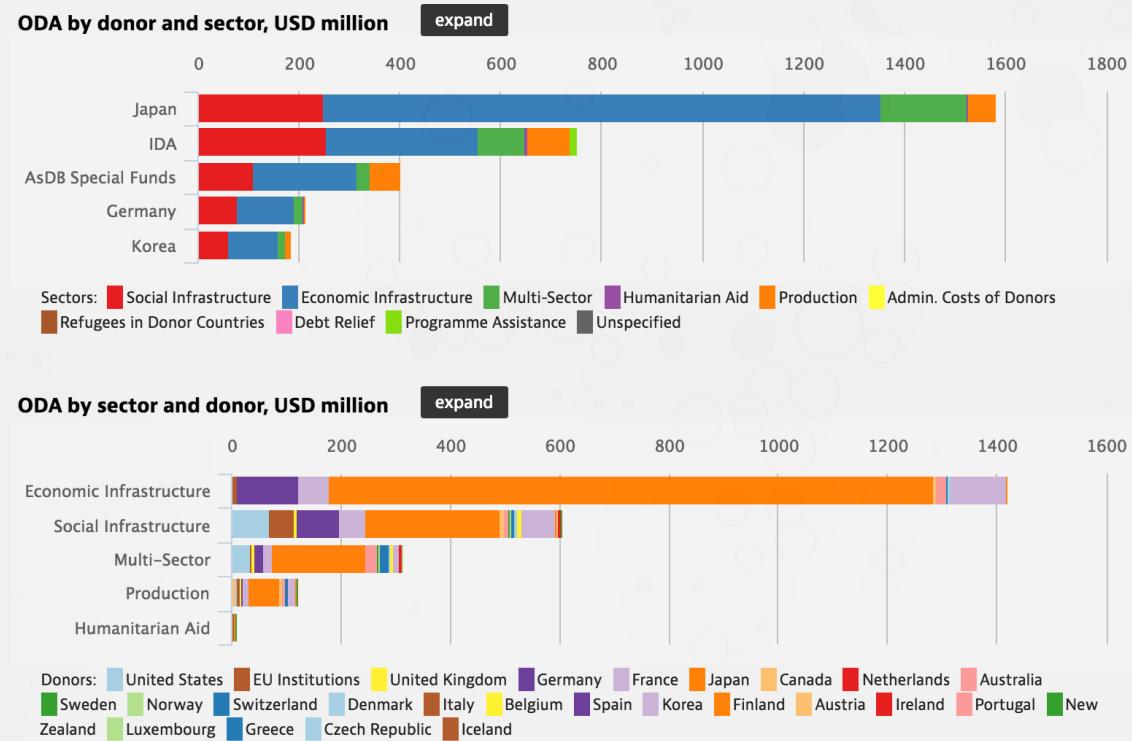
<http://www.oecd.org/dac/financing-sustainable-development/development-finance-data/aid-at-a-glance.htm>

¹⁷ Source: “Compare your country: Aid statistics by donor, recipient, and sector”

<http://www2.compareyourcountry.org/aid-statistics?cr=302&cr1=oecd&lg=en&page=1>

Viet Nam - Country profile

These charts present **gross disbursements of Official Development Assistance (ODA)** in 2016 from DAC donors and from multilateral organisations. You can expand each of the charts with the donor/sector breakdown to view more details and access a list of projects by donor and sector, realized in a given recipient country.



Notice from the chart that the largest donor country to Vietnam was Japan, followed by “IDA”, or the International Development Association, the part of the **World Bank** that helps the world’s poorest countries. According to the World Bank,

The IDA aims to reduce poverty by providing loans (called “credits”) and grants for programs that boost economic growth, reduce inequalities, and improve people’s living conditions.

IDA is one of the largest sources of assistance for the world’s 75 poorest countries, 39 of which are in Africa, and is the single largest source of donor funds for basic social services in these countries.

IDA lends money on concessional terms. This means that IDA credits have a zero or very low interest charge and repayments are stretched over 25 to 40 years, including a 5- to 10-year grace period. IDA also provides grants to countries at risk of debt distress.

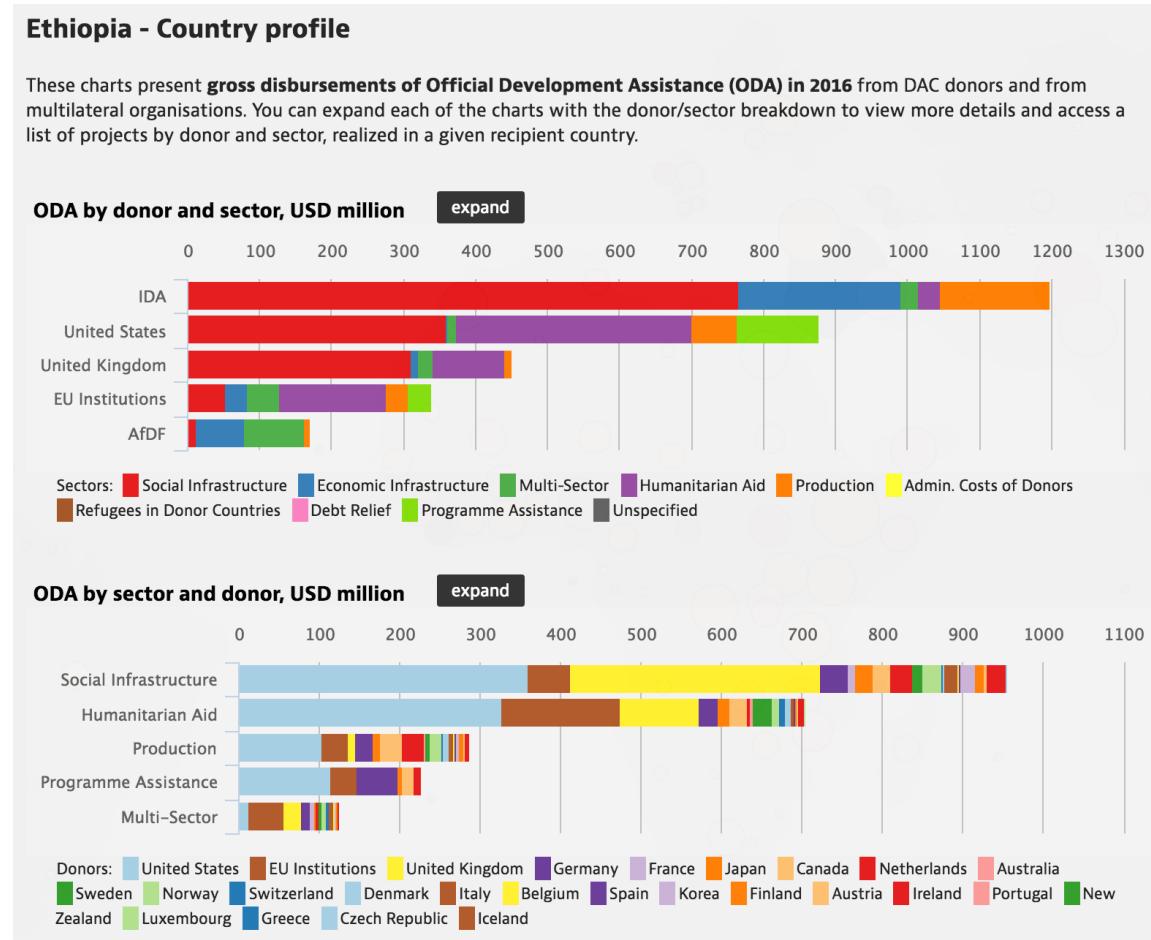
The colors in the top half of the chart represent the sectors within Vietnam that received aid, including “economic infrastructure” (the roads, bridges, transportation hubs, communication

technologies, and other physical infrastructure that makes a country's economy function efficiently) and "social infrastructure" (which includes schools, universities, hospitals, housing, and other goods that improve social welfare).

As we can see from the bottom of the chart, humanitarian aid makes up only a tiny proportion of Vietnam's total aid, which is almost entirely in the form of development aid.

Foreign aid case study #2 - Ethiopia

In 2016 Ethiopia was the world's fifth largest recipient of official development assistance, receiving \$1.98 billion.¹⁸ Below we can see the extent, nature, and sources of aid received by Ethiopia.



Like Vietnam, the IDA wing of the World Bank is a major donor to Ethiopia, providing \$1.2 billion of its almost \$2 billion in total aid. Concessional World Bank loans are being used largely to provide social infrastructure to Ethiopia (schools, clinics, hospitals, etc), followed by economic infrastructure.

¹⁸ Source: "Aid at a glance charts" from the OECD

<http://www.oecd.org/dac/financing-sustainable-development/development-finance-data/aid-at-a-glance.htm>

Following the IDA is the United States, which unilaterally provided Ethiopia with almost \$900 million in aid, much of which was in the form of humanitarian aid.

Notice also that compared to Vietnam, which received almost no humanitarian aid, over a third of the aid received by Ethiopia (\$700 million) was humanitarian. The difference can be explained by the fact that Vietnam is a relatively rich country that is not as prone to humanitarian emergencies like drought, conflict, flooding, and outbreaks of disease as compared to Ethiopia.

Evaluation of foreign aid

- Evaluate the effectiveness of foreign aid in contributing to economic development.
- Compare and contrast the roles of aid and trade in economic development.

There has been much debate over the effectiveness of aid in contributing to economic development. On one side of the discussion are economists like Columbia University's Jeffrey Sachs, who argued in a January 2014 column in *Foreign Policy Journal* that "the recent evidence shows that development aid, when properly designed and delivered, works, saving the lives of the poor and helping to promote economic growth." As evidence, Sachs points to the reduction in instances of malaria in sub-Saharan Africa since the turn of the 21st century¹⁹:

At the turn of the new century, malaria was front and center of the global aid debate. Research by myself and others, and evidence garnered in the report of the World Health Organization (WHO) Commission on Macroeconomics and Health that I had the honor to chair, showed that in addition to being a health catastrophe, malaria imposes a significant economic burden, particularly in sub-Saharan Africa. Luckily, though, the world was starting to take notice. In 2000, the U.N. Millennium Declaration, The African Summit on Malaria, and the G8 Declaration all addressed the burden of malaria and committed the world to action. The debate soon turned to the issue of policy: how could the malaria burden be reduced?

The combination of bed nets and effective medicines, supported by rapid diagnosis of infections, makes for a powerful one-two punch in saving lives and reducing malaria transmission.

Without financial support, poor people could not afford either the (bed nets) or the new medicines. Attempts to sell the nets at a discount, known as social marketing, had very little take up, since many poor families simply lacked any cash income at all. The prospect of achieving "mass action" protection through social marketing was very small. Moreover, impoverished households would often scrape together the needed money only to buy the cheaper but ineffective nets, rather than the more

¹⁹ Source: Sachs, Jeffrey. "The Case for Aid" Foreign Policy, January 21, 2014
<https://foreignpolicy.com/2014/01/21/the-case-for-aid/>

expensive but more effective (ones).

Governments of low-income African countries needed donor support for the scale-up effort since their own domestic tax revenues, even when amply allocated to public health, could not cover the costs of a basic primary health system including scaled-up malaria control. The financial calculations, laid out by the Commission on Macroeconomics and Health, showed that an impoverished country with a GDP of around \$500 per capita, typical for a poor country in Africa, may be able to muster around \$15 per person per year out of domestic revenues for primary health, while the costs of a basic public health system would be around \$50-\$60 per person per year.

... The WHO (World Health Organization) swung its powerful weight behind the mass free distributions of bed nets throughout sub-Saharan Africa. Soon after, U.N. Secretary-General Ban Ki-moon established the mass free distribution of bed nets as policy for all U.N. agencies, and called on the world's governments and NGOs to support the scale-up effort. Ban's leadership tipped the global scales decisively. Close to 300 million bed nets were freely distributed from 2008-2010, with the Global Fund to Fight AIDS, Tuberculosis and Malaria and the U.S. President's Malaria Initiative program paying for a substantial share of the scale-up.

The evidence is overwhelming that malaria declined precipitously as a result of these bold measures. WHO's latest report finds a stunning 51 percent drop in malaria deaths of African children under the age of five between the years 2000 and 2012. These results are historic. Roughly a half-million children, if not more, are being saved each year that otherwise would have succumbed to malaria.

According to Sachs and other aid proponents, without the work of both NGOs and official development agencies and the governments that support them, the poverty reduction and improvements in health and education experienced across the developing world in recent decades would have been significantly less pronounced.

On the other side of the debate are economists like New York University's William Easterly, who argues that due to factors like a lack of accountability and feedback, corruption, and the fact that recipient countries can become dependent upon and beholden to donor countries, rather than developing their own capacities as contributors to global trade, aid has failed to achieve meaningful improvements in the development of poor countries.

According to Easterly²⁰,

Economic development happens, not through aid, but through the homegrown efforts of entrepreneurs and social and political reformers. While the West was

²⁰ Source: Easterly, William "Why aid doesn't work?" CATO Unbound, April 2, 2006
<https://www.cato-unbound.org/2006/04/02/william-easterly/why-doesnt-aid-work>

agonizing over a few tens of billion dollars in aid, the citizens of India and China raised their own incomes by \$715 billion by their own efforts in free markets. Once aid agencies realize that aid CANNOT achieve general economic and political development, they could start concentrating on fixing the system that fails to get 12-cent medicines to malaria victims.

The two key elements necessary to make aid work, and the absence of which has been fatal to aid's effectiveness in the past, are FEEDBACK and ACCOUNTABILITY. The needs of the rich get met through feedback and accountability. Consumers tell the firm "this product is worth the price" by buying the product, or decide the product is worthless and return it to the store. Voters tell their elected representatives that "these public services are bad" and the politician tries to fix the problem.

Of course, feedback only works if somebody listens. Profit-seeking firms make a product they find to be in high demand, but they also take responsibility for the product—if the product poisons the customer, they are liable, or at least they go out of business. Elected representatives take responsibility for the quality of public services. If something goes wrong, they pay politically, perhaps by losing office. If it succeeds, they get the political rewards.

Aid agencies can be held accountable for specific tasks, rather than the weak incentives that follow from collective responsibility of all aid agencies and recipient governments for those broad goals that depend on many other things besides aid agency effort. Examples of the latter include such unaccountable goals as the very fashionable campaign to achieve the UN Millennium Development Goals, or the sweeping goals of economic growth, government reform, and democracy for poor countries mentioned above. If a bureaucracy shares responsibilities with other agencies to achieve many different general goals that depend on many other things, then it is not accountable to its intended beneficiaries—the poor. No one aid agent is individually responsible for successfully achieving any one task in the current aid system. Without accountability, then the incentive for finding out what works is weak. True accountability would mean having an aid agency take responsibility for a specific, monitorable task to help the poor, whose outcome depends almost entirely on what the agency does. Then independent evaluation of how well the agency does the task will then create strong incentives for performance.

At the very least, opponents argue, aid has played a secondary role, while the primary driver of economic development has instead been the opening up of global markets to less economically developed countries, trade liberalization, foreign direct investment, and other market-oriented approaches to raising incomes and reducing poverty.

4.7 The Role of International Debt

Foreign debt and its consequences

- Outline the meaning of foreign debt and explain why countries borrow from foreign creditors.
- Explain that in some cases countries have become heavily indebted, requiring rescheduling of the debt payments and/or conditional assistance from international organizations, including the IMF and the World Bank.
- Explain why the servicing of international debt causes balance of payments problems and has an opportunity cost in terms of foregone spending on development objectives.
- Explain that the burden of debt has led to pressure to cancel the debt of heavily indebted countries.

In the macroeconomics section of this course we learned how countries accrue debt: when a government's budget is in deficit, it must borrow money to finance that deficit, adding to the national debt. Every year a country experiences a budget deficit, its national debt grows.

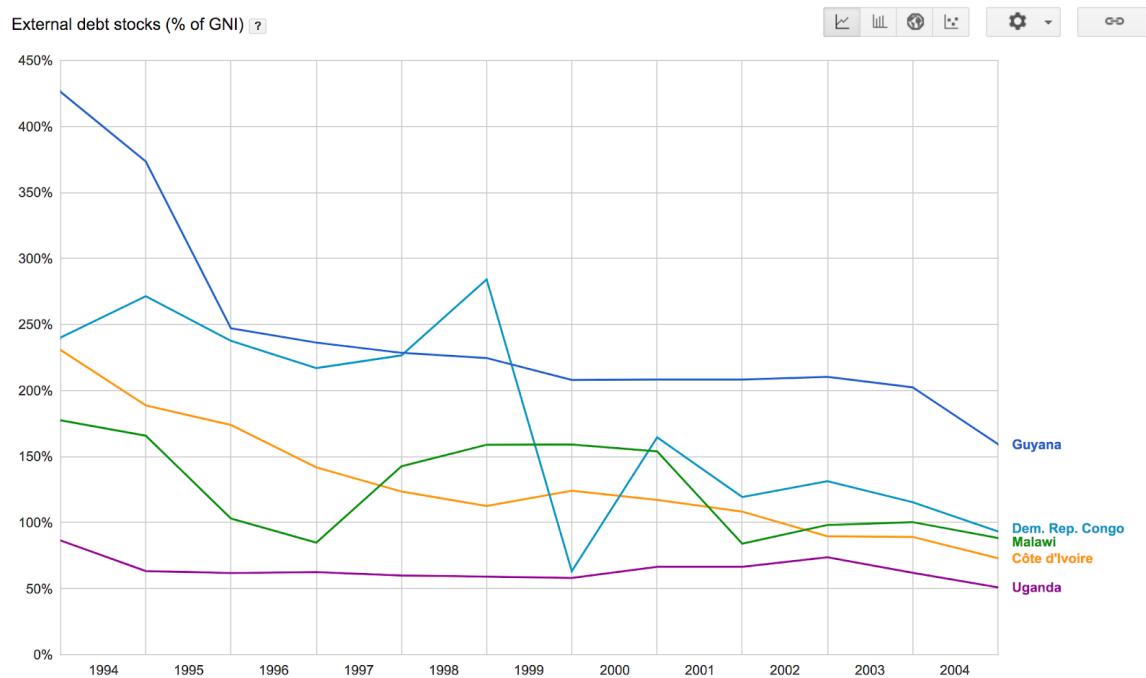
Debt in and of itself is not always a bad thing. The ability to borrow funds to finance a budget deficit allows a country to fund the current and capital expenditures necessary to keep the government running, to invest in the nation's infrastructure, and to provide the public goods required for economic development.

The countries with the highest debt levels in the world are hardly on the brink of economic collapse. The United States, with a national debt of over \$20 trillion (around 100% of GDP) enjoys macroeconomic stability and a strong currency. Japan has debt that equals nearly 250% of its national income, yet it is considered one of the safest economies in the world to do business in.

The problem arises when a country's debt is mostly owed to foreigners, AND when the burden of that debt limits the country's ability to provide necessary public goods and to invest in infrastructure. **External debt, or foreign debt**, is the proportion of a country's debt that is owed to international lenders, including commercial banks, governments, and international financial institutions like the World Bank and the IMF.

When a less economically developed country accumulates a large amount of foreign debt, servicing that debt can crowd out essential spending on public goods and infrastructure, and thus limit the level of economic development in the poor country. **Debt servicing** refers to the money that a government must spend to pay the interest on past debts. The larger a country's foreign debt, the more of its limited tax revenues it must allocate to service that debt. If the country has to borrow money to service past debts, its **total debt stock** will increase.

The chart below shows the total debt stocks of several LEDCs as a percentage of their GNIs from 1994 - 2004.

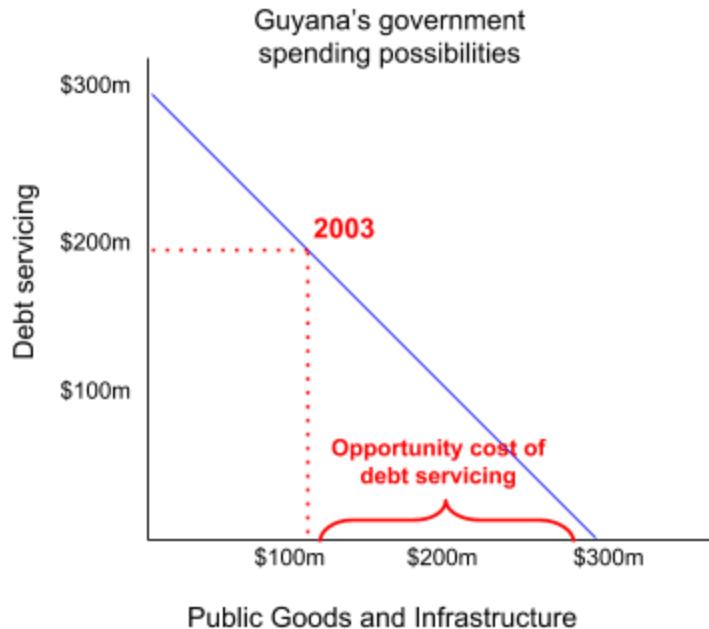


Notice that nearly all the countries averaged external debt stocks of well over 100% of their GNIs over the decade from 1994-2004.

To give an idea of the burden high levels of external debt can put on a country's economy, consider a country that must service external debt of 200% of GNI, such as Guyana in 2003.

- Assume Guyana owed external creditors 10% interest on its debt.
- With debt equal to 200% of GNI, this means that as much as 20% of Guyana's total income in 2003 was owed to foreign creditors in interest payments alone.
- Guyana's government must then collect taxes on its citizens of at least 20% of their total income, which would be just enough to service its debt.
- That 20% of income would then be handed over to foreign lenders, leaving Guyana with little or no money to spend on infrastructure, health, or education.

High levels of external debt can result in a poverty cycle or poverty trap in which foreign debt payments limit the ability of a country to achieve economic development, requiring the government to accrue more debt, leading to higher debt servicing costs. The dilemma of foreign debt can be illustrated in a simple PPC model showing the tradeoff between servicing debt and investing in public goods and infrastructure.



Assuming Guyana's GNI is \$1 billion in 2003 and that the government collects \$300 million in taxes. It must spend \$200 million on interest payments to foreigners to service its debt, leaving only \$100 million to invest in public goods and infrastructure. Guyanese households lose 30% of their income to taxes, but get only 10% back in the form of public goods, while 20% is "leaked" from the economy, ending up in the pockets of foreign creditors.

Servicing international debt also creates **balance of payments** (BoP) problems. Recall that the BoP consists of the current account (which measures the flow of money for the purchase of goods, income flows, and current transfers) and the financial account (which measures the flow for financial and real assets). Debt servicing leads to an outflow of funds in the financial account, moving it towards deficit.

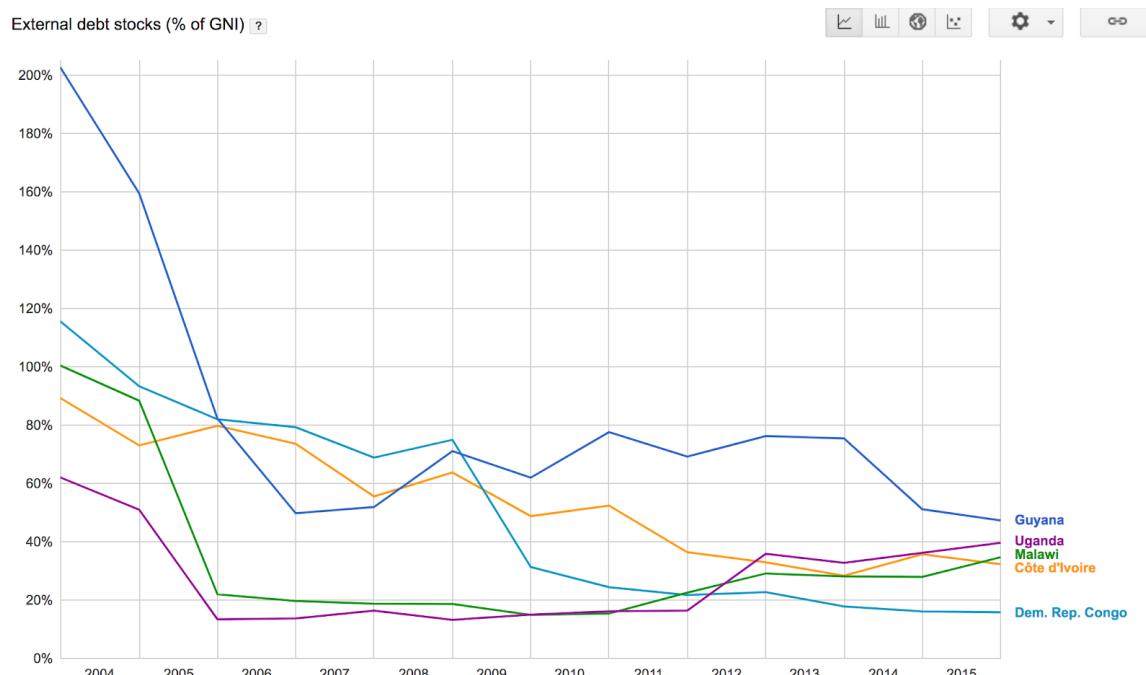
To pay back foreign creditors, a country must have foreign currency, which it can earn from selling exports (measured as a credit in the current account). However, if a country does not have exports to sell or if the demand for or the value of its exports suddenly falls (not uncommon in LEDCs that largely specialize in primary commodities), then the ability to service external its external debt is limited, and a country runs the risk of defaulting on its foreign debts. A default would result in a country being cut off from international credit markets and limit the government's ability to finance future budgets.

High levels of foreign debt pose an obstacle to economic development. For this reason, the international development institutions (the World Bank and the IMF) launched an initiative to reduce the foreign debts of the world's most heavily indebted poor countries (HIPC). According to the IMF²¹,

²¹ Source: "Debt Relief Under the Heavily Indebted Poor Countries (HIPC) Initiative" International Monetary Fund, <http://www.imf.org/>

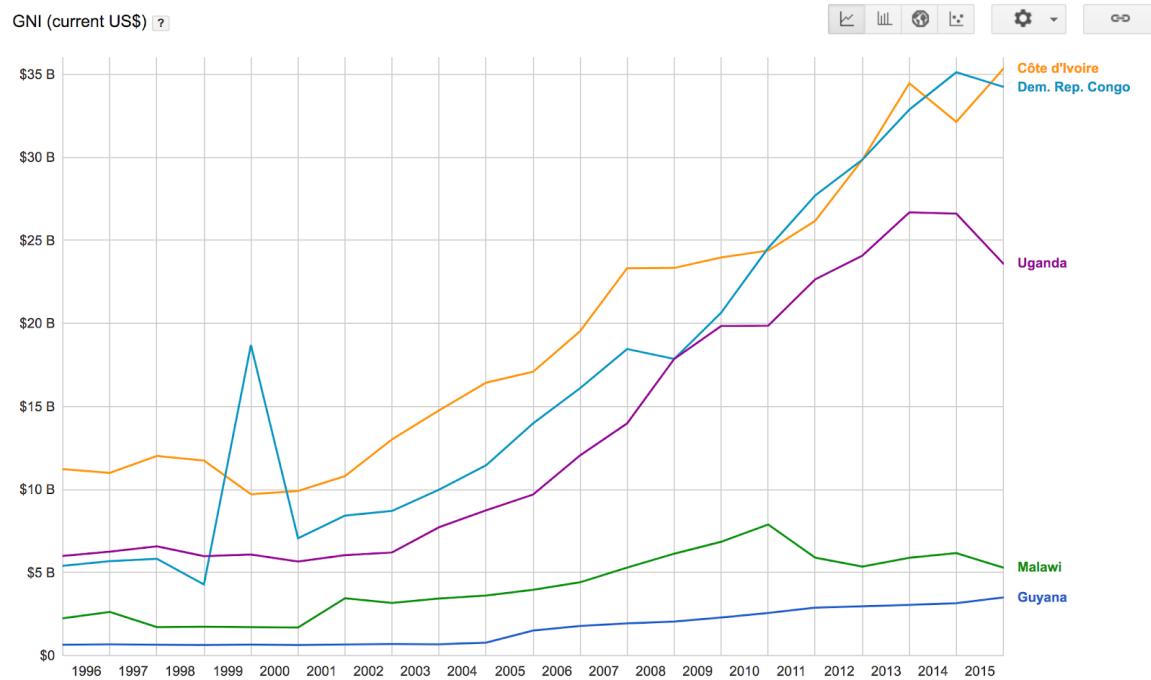
The HIPC Initiative was launched in 1996 by the IMF and World Bank, with the aim of ensuring that no poor country faces a debt burden it cannot manage. Since then, the international financial community, including multilateral organizations and governments, have worked together to reduce to sustainable levels the external debt burdens of the most heavily indebted poor countries.

Since the launch of the HIPC initiative, the debt levels of the 36 most indebted poor countries have fallen dramatically. Consider the five countries from our chart above. Below we can see what has happened to their debt levels since 2004.



Since 2004 the debt levels of all five countries have fallen, from more than 200% of GNI to less than 50% in Guyana, from 100% to under 40% in Malawi, and from 115% to under 20% in the DRC.

Among the 36 countries receiving debt relief under HIPC, the reduction in the debt servicing costs has freed up public funds to invest in infrastructure and public goods. The chart below shows how the five countries' incomes have changed since the launch of HIPC in 1996.



Debt relief helps countries escape the debt poverty cycle, freeing up scarce government resources for investments in infrastructure, health, and education.

Furthermore, the foreign revenues earned from exports can be used to buy much needed imported capital goods and technology, including medical equipment, drugs, communications and transportation technologies, and other goods that improve the living standards of the country's people, rather than on servicing external debt.

Debt relief reduces the chance of a sovereign default (i.e. government bankruptcy) and leads to greater exchange rate stability on the forex markets, as international investors are more confident in a poor country's ability to repay new debts if it has a smaller level of past debts.

In concert with domestic factors such as improvements in education and health, international factors such as diversification of exports and economic integration, foreign direct investment, and foreign aid, debt relief has played a crucial role in the economic development experienced by the world's poorest countries over the last couple of decades, and will likely continue to do so in the decades ahead.

4.8 The Balance Between Markets and Intervention

In this final section of the course we will revisit one of central themes of economics introduced way back at the beginning of the IB Economics course: the role of government in the economy.

In our study of microeconomics we examined how government intervention in markets can both reduce efficiency and welfare (think of rent controls, agricultural subsidies, and taxes on perfectly harmless goods), and increase efficiency and welfare (think of government provision of public goods, environmental protections, and taxes on harmful goods).

When looking at the role of government in the national economy, we debated the extent to which government should intervene to promote the macroeconomic objectives of full employment, price stability, and economic growth. The laissez faire view argued that government intervention can only make a country's economy less efficient and lead to instability in the business cycle. The Keynesian viewpoint, meanwhile, portrayed government as the great stabilizer: the captain that steers the economy through the turbulent and uncertain seas of macroeconomic volatility.

In the realm of international trade we considered the intent and effects of government protectionism, the impact of forex market intervention, and the role governments can play in promoting their countries' export sectors or in balancing their current and financial accounts.

In every area of our study of economics the balance between free markets and government intervention has been at the core of our analysis and evaluation. Likewise, in looking at economic development, we can once more consider the extent to which improvements in the living standards of people in poor countries are best achieved through the free market, through government intervention, or through a complementary approach, blending a balance between markets and intervention.

Strengths of market-oriented policies

- Discuss the positive outcomes of market-oriented policies (such as liberalized trade and capital flows, privatization and deregulation), including a more efficient allocation of resources and economic growth.

Market-oriented policies for economic development include trade liberalization, economic integration, privatization of state-run enterprises, deregulation, and foreign direct investment.

Policies driven by the free market have some strengths over government-driven development strategies. Advantages of different market-oriented policies include:

- **Of trade liberalization:** Freer trade stands to raise incomes in poor countries by

opening rich country markets up to their exports. Additionally, access to imported technology and capital stands to increase productivity in LEDCs and provide access to welfare improving goods like medicines and communications technologies.

- **Of economic integration:** When poor countries join trading blocs with their neighbors or with richer countries, a more efficient allocation of resources results as trade is “created” between member countries. There will be an overall increase in output between countries participating in trading blocs, and a corresponding increase in incomes.
- **Of privatization:** State-run enterprises like government-run mining and energy firms, airlines, utilities, and other large industries are notoriously inefficient as they are not driven by the profit-motive that leads privately firms to produce their goods in the least-cost manner. Privatization of state-run industries in LEDCs can benefit the residents of those countries as markets are opened up to competition, increasing product variety and quality and potentially driving down costs.
- **Of deregulation:** Overly burdensome regulations deter domestic and foreign investment and have the potential of stifling innovation and productivity. Deregulation (reducing the amount of government interference in how goods are produced) can therefore spur investment and increase competition and efficiency, leading to increased provision of the goods and services needed to improve life quality in poor countries.
- **Of foreign direct investment (FDI):** Many LEDCs lack the infrastructure and capital goods required for economic growth and development. By attracting FDI (when foreign firms take an ownership stake in domestic enterprises) poor countries can secure productivity-enhancing technologies that would otherwise be difficult to obtain.

Market-oriented policies are generally good at improving efficiency and boosting economic growth. Goods and services demanded by domestic and foreign consumers tend to be provided relatively efficiently by the free market. Markets fail, however, in many ways, leaving the door open for government intervention.

Weaknesses of market-oriented policies

- Discuss the negative outcomes of market-oriented strategies, including market failure, the development of a dual economy and income inequalities.

An exclusively market-oriented approach to economic development stands to fail at achieving long-term, sustainable improvements in human welfare. The weaknesses of market-oriented policies include:

- **Market failure:** Not all goods will be efficiently provided by the free market. Harmful effects of production, such as pollution and impacts on health, will be unregulated and could inhibit improvements in social welfare. Certain goods, like health and education, will be underprovided by the market, while others, such as security and sanitation, may not be provided at all. The market also fails to protect common access resources like fisheries and forests, so market-oriented policies tend to result in the unsustainable exploitation of certain valuable resources, particularly in

- poor countries that lack the voice or influence to stand up to profit-seeking multinational corporations.
- **The development of a dual economy:** When opening up to international trade, poor countries run the risk of developing a dual economy, in which a segment of the workforce and resources are allocated towards producing goods (usually primary commodities) for the global market while goods that actually benefit domestic consumers go underproduced. One sector is focused on local needs and another to the global market. Market-oriented policies may neglect the needs of locals at the expense of the demands from international stakeholders like MNCs.
 - **Income inequality:** There is no guarantee that the gains from market-driven growth will be equitably distributed across a poor country's population. In fact, without active government efforts to distribute the gains from economic growth across all of society, it is highly likely that income inequality will increase while a country opens itself up to the global market. Inequality can pose an obstacle to economic development as large swaths of society are denied the health and education benefits enjoyed by those at the top of the income ladder.

Strengths of interventionist policies

- Discuss the strengths of interventionist policies, including the provision of infrastructure, investment in human capital, the provision of a stable macroeconomic economy and the provision of a social safety net.

While market-oriented policies for growth and development involve little or no role for the government, **interventionist policies** are those that involve the government providing or supporting the provision of certain goods, services, programs, or institutions.

Government can fill the gaps left by free markets. For example, each of the following are areas where the free market either underprovides or fails altogether to provide something essential for economic development to take place:

- **Infrastructure:** Sanitation, transportation, energy, waste collection, communication, and other infrastructure goods, when left to the free market, will be underprovided in a country. Thus, it is left to the government to intervene in these markets to assure that the economic and social infrastructure necessary for development exists.
- **Macroeconomic stability:** Without government policymakers managing the level of aggregate demand through fiscal and monetary policies, the free market will experience swings in output, employment, and prices that can undermine sustainable growth and development.
- **Investments in human capital:** Throughout our study of economic development we have emphasized the importance of health and education in determining a people's living standards. As these merit goods, which provide benefits not just to those who consume them but to society as a whole, tend to be underprovided by the free market, government must take measures to assure all citizens have access to them.
- **Social safety nets:** Without programs like unemployment insurance, the welfare system, and old-age pensions, the benefits of the market-oriented economic growth

will be enjoyed by relatively few in society. Social safety nets prevent those who lose jobs, grow old, or fall into poverty from slipping through the cracks and help spread the gains from economic growth across all of society.

Weaknesses of interventionist policies

- Discuss the limitations of interventionist policies, including excessive bureaucracy, poor planning and corruption.

Too much of anything can be a bad thing; and this of course holds true for government intervention. An overly bureaucratic state can limit investment or create an environment of uncertainty and corruption. When too much of the economy is overseen or managed by the government, decisions will be made that result in a poor or inefficient allocation of resources, limiting productivity and income gains and slowing the rate of economic growth.

Poor countries that adopt a development strategy that places total emphasis on government planning and production will not enjoy the great gains brought by trade liberalization, increased competition, privatization, foreign investment, and the other market-oriented policies.

The “golden mean” for economic development, it seems, is a balance between markets and intervention.

Market with government intervention

- Explain the importance of good governance in the development process.
- Discuss the view that economic development may best be achieved through a complementary approach, involving a balance of market oriented policies and government intervention.

The key to economic development for the world’s poorest countries today lies in good governance. According to development economist Paul Collier²²,

The relationship between the quality of governance and national economic performance is highly dependent on circumstances. When the conditions are right, economies can grow fast despite poor governance (e.g. Bangladesh). But poor governance is generally bad for the economy, and really bad governance can destroy economies (e.g. Zimbabwe).

Governance is often bad in poor countries because their leaders, many of whom are among the ‘global super-rich’, benefit enormously from their roles as gatekeepers between their national economies and the international system and it pays to keep their citizens uneducated and ill-informed.

²² Source: Moore, Mick “Improving Governance in Bottom Billion Countries” Institute of Development Studies, March 2008 <http://www.ids.ac.uk/files/NewNo9-Governance-web.pdf>

A corrupt government can be the ultimate obstacle to economic development. Therefore, in order for any of the strategies (market-oriented, interventionist) outlined in this section of the course to be effective, the government must have the wellbeing of its citizens as its first priority.

Effective governance requires that officials in power be accountable to the citizens; usually this means they are democratically elected and can be pushed out of power through the ballot box if their actions do not adequately meet the demands of the voting public.

However, economic development has also been achieved by non-democratic countries, such as China, through heavy emphasis on interventionist policies combined with a dose of market-oriented growth.

In the end, meaningful improvements in human welfare, including longer lives, greater access to education, human rights, employment, gender and racial equality, and many other indicators of human development, will be achieved through a complementary approach, involving a balance of market-oriented policies and government intervention.