ECON 103: Principles of Macroeconomics Lecture Notes

Aina Katsikas, M.S.

Department of Economics, University of Nevada, Reno

Summer 2022: Second Session

Course Outline

These lecture notes are adapted from Macroeconomics by Acemoglu, Laibson & List 3rd edition.

1. Week 1
    - Principles of Economics
        - The Scope of Economics
        - Three Principles of Economics
        - Optimization
        - Equilibrium
        - Empiricism
    - Economic Science
        - The Scientific Method
        - Causation and Correlation
2. Week 2
    - Optimization
        - Optimization Using Total Value
        - Optimization Using Marginal Analysis
    - Demand, Supply and Equilibrium
        - Markets
        - Buyer Behavior
        - Seller Behavior
        - Supply and Demand in Equilibrium

3. Week 3
    - Macroeconomic Aggregate Metrics
        - Macroeconomic Questions
        - National Income Accounts
        - What is not included in the GDP?
        - Real vs. Nominal
    - Aggregate Incomes
        - Inequality Around the World
        - Productivity and the Aggregate Production Function
        - The Role and Determinants of Technology
4. Week 4
    - Economic Growth
        - The Power of Economic Growth
        - What fuels Economic Growth?
        - The Malthusian Model
        - The Solow Growth Model
    - World Development
        - Proximate vs. Fundamental Causes of Prosperity
        - Institutions and Economic Development
5. Week 5
    - Employment & Unemployment
        - Measuring Employment and Unemployment
        - Equilibrium in the Labor Market
        - Voluntary and Frictional Unemployment
        - Wage Rigidity and Structural Unemployment
        - Cyclical Unemployment and the Natural Rate of Unemployment

1 Week 1

1.1 Principles of Economics

Economics is about so much more than money. Economists study all human behaviors and the choices we make. Choices, not money, is the unifying feature of all economic studies. Economists think of all human behavior as the outcome of choices. For example, a dad tells his teenage daughter to wash the family car. She has a few options. She can wash it, or negotiate for an easier chore,

or refuse to wash it and suffer the consequences. Everything you do is the outcome of choice you are making.

An **economic agent** is an individual or a group that makes choices. For example, a consumer can choose which items to purchase in a grocery store. A citizen can choose whether or not to vote. A parent can choose how to reward their kids. A government, firm, labor union, or even a street gang, are all examples of economic agents. These groups make decisions.

Economics also studies the allocation of scarce resources. **Scarce resources** are things that people want, where the quantity that people want exceeds the quantity that is available. For example, iPhones, ice cream, gold wedding bands, are all examples of scarce resources. A resource does not necessarily need to be a luxury good to be scarce. Everyday consumer goods are also scarce, like toilet paper and clean drinking water. Scarcity exists because people have unlimited wants in a world of limited resources. The world does not have enough resources to give everyone everything they want for free. Economics studies how these scarce resources are distributed across systems.

How are scarce resources allocated? They go to the consumers who are able and willing to pay for them. This explains why we have to pay a price for the goods we consume, because almost all goods utilize scarce resources. Each consumer has 24 hours in a day, which can be distributed to various tasks. One of which is labor, which can produce wages for us to purchase items. We can also dedicate our time to leisure, which can include hobbies or free time. You choose how many of those 24 hours you will allocate to various activities. We can use economic reasoning to compare the costs and benefits of alternative options to make the best choice.

Economics is the study of how agents choose to allocate scarce resources and how those choices affect society. Choices play a key role in the formal definition of economics. However, there is another dimension to economics and that includes the effects of any individual's choices on society. For example, the sale of a new car does not just affect the person who purchased it. The sale generates a sales tax, which is used by the government to purchase various things. That new car also generates more congestion on public roadways. That car also produces wages for the sales person that sold it to you. Economists study the original choice and it's multiple consequences for other people in the world.

Positive Economics and Normative Economics

There are two broad categories of economics, and that is positive and normative economics.

Positive economics describes what people actually do. It is the analysis that generates objective descriptions or predictions, which can then be verified with data. For example, from March to April 2020, the percentage of the US labor force that was unemployed increased from 4.4% to 14.7%. Describing with what has happened or predicting what will happen is all under the umbrella of positive economics. Predictions are considered positive economics because they can either be confirmed or refuted with the use of future data collection.

Normative economic analysis includes recommendations of what an individual or society should do. Most normative economics relies on subjective judgments. For example, consider an economic consultant providing guidance to an individual looking to invest her savings. The economist might ask about her preferences regarding risk. The economist would then explain that lower risk investments come at a cost, and that is a lower average rate of return in the long run. If she prioritizes stability over returns, then the economist would recommend low risk investments for her, and would construct a portfolio that delivers the level of risk that the client wants. When economic analysis is used to help individual economic agents choose what is in their personal best interest, this type of normative economics is prescriptive economics. Normative analysis generates advice to societies in general. For example, it informs public policy debates For government entities. When a government policy creates winners and losers, economists need to conduct normative analysis and make ethical judgments when evaluating policies. For example, is it ethical to create environmental regulations that prevent a real estate developer from draining a swamp so she can build new homes? What if those regulations protect migratory birds that other people value? Are there possible compromises? These public policy questions ask what society should do.

Another distinction in economics is between the two fields of microeconomics and macroeconomics. Microeconomics is the study of how individuals, household, firms, and governments make choices and how those choices affect prices, the allocation of resources, and the well-being of other agents. For example, the study of how coal-based pollution affects healthcare expenditures is an example of microeconomics. In general, microeconomists evaluate a particular piece of the overall economy. Macroeconomics is the study of the economy as a whole. Macroeconomists study economy- wide phenomenon, like the growth rate of a country's total economic output, the inflation rate, or the unemployment rate. A good example of macroeconomics is the percentage increase in overall prices, such as inflation, or the fraction of the nation's labor force that is looking for work but cannot find a job (the unemployment rate.) Macroeconomists search for the best policies to stimulate an economy during a recession.

Three Principles of Economics

There are three foundational principles of economics. Those include:

1. **Optimization**

   Optimization means trying to pick the best feasible option, given limited information, knowledge, experience and training. Economists do not believe that people succeed in picking the best feasible option, however they believe that people try to pick the best feasible option. Optimization explains the choices that people make, including minor and major decisions.

2. **Equilibrium**

   Equilibrium is a state in which everyone is simultaneously optimizing. Nobody believes they would benefit personally by changing their own behavior, given the choices of others. For example, equilibrium in the labor market is when the quantity of people who want to work is equal to the quantity of job positions open for hire.

3. **Empiricism**

   Empiricism is analysis that uses data to develop theories, test these theories, evaluate the success of different government policies, and determine what is causing things to happen in the world. Analysis uses data, this is empiricism.

Now, I will dive into more detail on these three concepts.

Optimization

Economists believe that people optimize, meaning that economic agents try to choose the best feasible option, given whatever limited information or knowledge they have. Feasible options include those that are available and affordable. For example, if you have $10 in your wallet, then a $60 target shopping spree is not feasible. Economists like to debate on the concept of feasibility. For example, feasibility constraints may go beyond simple financial budgets. Optimization means that you weigh the information that you have, knowing that you cannot perfectly foresee the future. When an individual chooses the best feasible option given the information that is available, that decision maker is being rational, or exhibiting rationality. Rational thinking just requires a logical appraisal of costs, benefits, and risks known to the economic agent. Evaluating the rationality of a decision requires examining the quality of your initial decision, not the outcome.

In cases where economic agents consistently fail to choose the best feasible option, we will discuss normative economic analysis to help them realize their mistakes and make better choices in the future.

It is important to note that not everyone optimizes the same objective. For example, not every household is aiming to maximize their personal income. Most people aim to optimize their overall well-being, which consists of a mixture of health, wealth, and leisure. Governments for example, try to optimize a variety of things. This can include a mix of policy goals, like economic output in a population health. The COVID-19 crisis is a good example of the complexity of optimization. If a government aimed at maximizing economic output, it would not have shut down the economy via lockdown measures. Instead, it aimed to optimize public health. By optimizing public health, it sacrificed economic output. Optimization does not mean that everyone wins. There are sacrifices that have to be made in order to obtain certain objectives. In other words, there are trade-offs to our decisions. Budget constraints guide these optimization decisions. They highlight what we can afford, the cost of our choices, and which decision is feasible.

All optimization problems include trade-offs. An economic agent faces a trade-off when the agent needs to give up one thing to get something else. For example, if you spend one hour exercising, you cannot spend that hour earning a wage at your job. Individual consumers face a trade off of work over leisure. Instead of working, we can enjoy our leisure time. Instead of leisure time, we could choose to work. Economists use budget constraints to describe trade-offs. A budget constraint shows the goods or services that a consumer can choose given her limited budget. A budget does not necessarily have to pertain to financial means. A budget can extend to the idea of time capacity. For example, we each have 24 hours in a day. That means, we are constrained within these 24 hours to choose which activity to spend our time on. If we had to choose between

leisure and work, there are many combinations of how we can share our time. For example, we could spend 12 hours on leisure and 12 hours on work. Or we could spend 20 hours on leisure and only four hours on work. The combination you choose depends on your preference as an economic agent.

When agents choose how to allocate their time, they face trade offs. In order to spend their time on one activity, they must give up an alternative activity. For example, as I am writing these lecture notes, I am giving up time to do a variety of other activities including jogging, reading, watching Netflix, and walking the dog. Out of all those alternatives, the best option for me is probably jogging because that is my favorite hobby. Jogging, therefore, is the opportunity cost in this case. Opportunity cost is the next best alternative use of a resource. In this case, the resource is my time. This is what an optimizer is effectively giving up when she allocates an hour of her time. The concept of an opportunity cost applies to all trade-offs, not just your daily time budget of 24 hours. For example, suppose that a factory could produce one of the two: 1 desk or 1 dining table. Let's say they both take the same amount of time and resources, yet the factory chooses to produce 1 dining table. The opportunity cost in this case is 1 desk. This is because the factory sacrificed 1 desk in order to produce 1 dining table. Therefore, the opportunity cost of producing 1 dining table is 1 desk.

To extend our evaluation, we can perform a cost benefit analysis and attach a monetary value to an opportunity cost. A cost benefit analysis is a calculation that identifies the best alternative, by summing benefits and subtracting costs, with both benefits and costs denominated in a common unit of measurement, like dollars. A cost benefit analysis is used to identify the alternative that has the greatest net benefit. The net benefit is the sum of the benefits of choosing an alternative minus the sum of the costs of choosing that alternative. For example, we can conduct a cost benefit analysis to decide whether we should fly to Miami or drive to Miami instead. If we are traveling with a friend, the cost of gas is $200 each. This is in comparison to a plane ticket of $300 per person. At first glance, driving is the cheaper option. However, we need to take into account the additional travel time required with driving. Driving takes 50 hours compared to flying, which would only take 10 hours. The cost of our time can be determined as the hourly wage that we could have earned if we had worked instead of spending that time on travel. For example, let's say our hourly wage is $13 an hour. Using this number, we cannot compute the net benefit for each of the two travel options.

Benefits - Costs = Net Benefits

The benefits in this case include the $100 in savings if we choose the driving route. The cost in this case is the extra 40 hours of travel time that cost us approximately $13 an hour. We can compute the net benefit of driving below.

$$NetBenefit = \$100 - (40 * \$13)$$

$$NetBenefit = \$100 - (\$520)$$

$$NetBenefit = -\$420$$

Therefore, the net benefit of driving is a negative value of $420. This means that driving costs us an extra $420 compared to flying, and that's after taking into account the cost savings of gas and the expensive cost of our time.

Hence, an optimizer would choose to fly instead of drive. Cost benefit analysis is an example of optimization. When you pick the option with the greatest value in net benefits, you are optimizing. Cost benefit analysis is useful for normative economic analysis. It determines what individuals in a society should be doing. It yields useful positive economic insights.

Equilibrium

It is important to understand that you are not the only economic agent practicing optimization. Each one of us is practicing optimization. Equilibrium is a special situation in which everyone is trying to optimize, therefore nobody believes they would benefit personally by changing his or her own behavior. A good example of this is demonstrated in the checkout lines at a grocery store. Let's say there are four different checkout lines, some lines may be longer or shorter than others. As a consumer, you would naturally prefer the shorter lines to save your time and resources. If you feel the need to switch to a shorter line, it is because you are practicing optimization. Therefore, because you feel like you would benefit by changing your behavior, the checkout line market is not in equilibrium. However, once all of the checkout lines are balanced and it would not benefit you to switch lines, then we can say that the market is in equilibrium. For example, look at the image comparisons below to understand the concept of equilibrium.

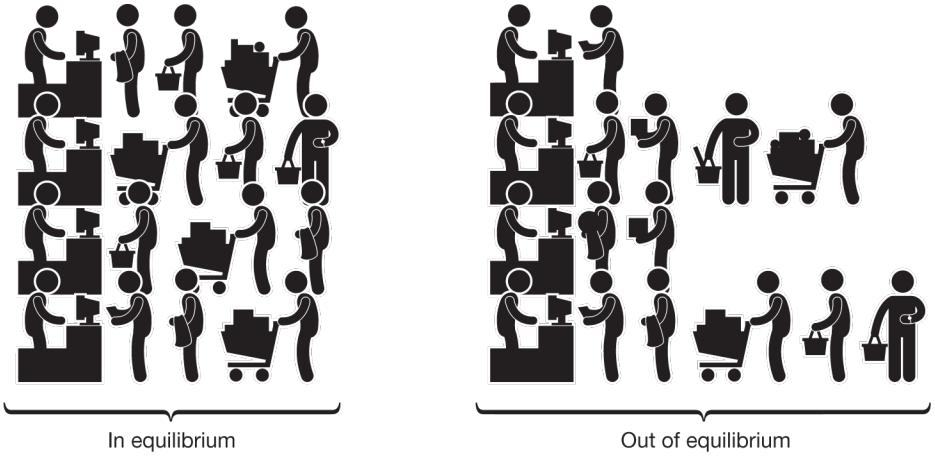

Figure 1.1: Equilibrium

When each line is the same length, the market is in equilibrium. Not a single person would benefit by switching lines.

If we are looking at the market for iPhones, the equilibrium points would be where the number of iPhones produced equals the number of iPhones sold. At this point, Apple is happy with the

fact that each of their products are being sold and they do not have to store any surplus. From the consumer standpoint, every consumer is happy with the fact that they can obtain the iPhone that they want.

The broad idea of equilibrium is that the number of consumers that want to purchase a product are able to purchase that product because the producer manufactures and sells a sufficient supply. For example, a shortage of toilet paper would be an example of a disequilibrium. In this case, the quantity demanded exceeds the quantity supplied. The number of toilet rolls in demand is much higher than the number of toilet rolls available for sale on the shelf.

Equilibrium analysis helps to evaluate, explain, and prevent the free rider problem. In the case of a free rider problem, and individual benefits from the work of others without contributing any work themselves. This can be problematic in societies. Equilibrium analysis helps to explain why individuals sometimes fail to serve the interests of society and how the incentive structure can be re-designed to fix these problems.

Empiricism

Empiricism is the idea that economists use real data to validate or refute their theories. In other words, economists make hypotheses about certain choices, human behaviors and economic outcomes. We use data to then confirm or deny the truth of our hypotheses. Empiricism is important because we need to prove our theories with evidence. It is not possible to prove a theory with more theory.

1.2 Economic Science

The Scientific Method

Economist work to evaluate the past and make forecasts about the future. Empiricism is at the heart of all of this. The choices that people make provide the data that we analyze to then make forecasts and predictions. Empiricism is also at the heart of scientific analysis in the scientific method. The scientific method is the name for the ongoing process that economists and other scientists use to: 1) develop models of the world and 2) evaluate those models by testing them with data. Given the complexity of the world, we do not expect the scientific method to generate a perfect model of economic behavior. We will never be able to precisely predict the future. However, we can partially predict the future.

Models and Data

Economists use data to develop models. A model is a simplified description of reality. Sometimes economists will refer to a model as a theory. These terms can be used interchangeably. Because models are simplified, they are not perfect replicas of reality. For example, maps are a good example of models. They are simplified versions of reality that help us navigate our selves through cities in the world.

Scientific models are used to make predictions that can be checked with empirical evidence. These are facts that are obtained through observation and measurement. We also refer to empirical evidence as data. A model's predictions are referred to as hypotheses. Whenever hypotheses are contradicted by the available data, economists return to the drawing board to create a better model that yields new hypotheses.

Let's look at an economic model for an example. All models begin with assumptions. Let's assume that each additional year of education increases an individuals annual earnings by 10%. Based on this assumption, we can develop a model to predict an individual's hourly wage if they pursue an extra year of education. For example, if someone earns \$15 per hour given that they have 13 years of education, let's predict what they would earn if they had 14 years of education.

$$1.10 * \$15 = \$16.50$$

Therefore, an additional year of education will increase their hourly wage to \$16.50 an hour compared to the previous wage of \$15 per hour. We can also use this same model to predict how 4 additional years of education will have an effect on increased earnings.

$$1.10 * 1.10 * 1.10 * 1.10 = 1.10^4 = 1.46$$

This implies that an additional four years of education will produce a 46% increase in wages.

This is the return to education model. It predicts the returns in annual earnings from an additional year of education. Most economic models are much more complex than this. In some models, it takes pages of mathematical analysis to derive the implications of the assumptions.

There are two important properties of models:

1. Models are only approximations.
2. A model makes predictions that can be tested with data.

It is important to remember that a model is only an approximation. Although an additional year of education does produce more earnings, the estimate of 10% may be slightly inaccurate. For example, an extra year of education in a chemistry degree versus economics degree might produce different levels of additional earnings. Another thing to remember is that a model makes predictions that can be tested with data. In this case, the data is collected from people's educations and earnings.

There are two mathematical definitions to understand before performing an economic analysis, the mean and median. The mean is the sum of all the different values in a data set divided by

the number of values. The median is the value calculated by ordering the numbers from least to greatest and then finding the value halfway through the list. I have examples below.

Let's say we are given a data set of wages for five people:

1. Person 1: $10 per hour
2. Person 2: $12 per hour
3. Person 3: $20 per hour
4. Person 4: $9 per hour
5. Person 5: $25 per hour

We can compute the mean with the computations below:

$$Mean = \frac{10 + 12 + 20 + 9 + 25}{5}$$

$$Mean = \frac{76}{5}$$

$$Mean = \$15.2$$

We can also compute the median by ordering the numbers from smallest to largest, and selecting the value in the middle.

9 10 12 20 25

Therefore, the median value is $12 per hour. The median seems like a simple metric, however it is a good snapshot summary of a data set. Essentially, 50% of the data set includes values above the median and 50% of the data set includes values below the median. Another cool thing about the median is that it is not affected by extreme values. For example, the mean value is $15.2 an hour yet the median value is $12 per hour. The mean value is higher because it is affected by the extreme values of $25 and $9.

Causation and Correlation

People often mistake correlation for causation. Causation occurs when one thing directly affects another. Causation is the path from cause to effect. For example, turning on the sprinklers causes the grass to get wet. A changing factor or characteristic is a variable. For example, the temperature of water in a tea kettle is a variable that fluctuates when you turn the stove on. Causation occurs when one variable directly causes another. For example, turning the sprinklers on directly it causes the grass to get wet.

Correlation means that two variables tend to change at the same time. As one variable changes, the other changes as well. It might be cause-and-effect, however correlation can occur when causation is not present. It is possible for two things to be correlated and yet not demonstrate any causation. For example, there is evidence that students who take music classes tend to score higher on the SAT. One may mistake this correlation for a causation and claim that music classes cause higher SAT scores. However, this is not the case. Students who take music lessons 10 to have with your parents who can pay for tutors to raise their kids SAT scores. In this case, parental wealth causes music lessons and SAT tutors and hire SAT scores. Music lessons are not causing higher SAT scores. When two variables are correlated, it suggests that causation may be present. However, further investigation is needed to validate the claim of causation.

Correlation is divided into three categories:

1. Positive correlation implies that two variables move in the same direction. For example, ice cream sales and summer temperatures are positively correlated. As ice cream sales increase, so do summer temperatures. Another example of this is that a decrease in public education spending is correlated with a decrease in high school graduation rates.

2. Negative correlation implies that two variables tend to move in opposite directions. For example, as housing prices increase in an area, the number of people immigrating to that area reduces.

3. Zero correlation is all the variables that have movements that are not related. For example, the number of friends you have is completely unrelated to the amount of vegetables you eat in a week.

There are two potential explanations for why correlation does not directly imply causality:

1. Omitted variables

   An omitted variable is something that has been left out of a model. If it were to be included, it would explain why two variables in a study are correlated. For example, an economist looking to predict an individual's annual income based on their number of years of education would find a high correlation. However, education might not be directly causing higher income because there are several variables left out of this equation. For example, age, health and area of expertise are all important factors that contribute to one's annual income.

2. Reverse causality

   Reverse causality occurs when we mix up the direction of cause-and-effect. For example, there is a high correlation between health and wealth. Healthy people tend to be wealthy, and wealthy people tend to be healthy. It is incorrect to say that good health causes wealth, because it is also possible that wealth causes good health. Wealthy individuals have access to that are healthcare, resources, and clean living environments compared to lower income individuals.

One method of determining cause and effect is to run an experiment. An experiment is a controlled method of investigating causal relationships among variables. For example, psychology centers run experiments on patients to make a new discoveries in the world of behavioral health.

Important component of every experiment is the concept of randomization. Randomization is the assignment of subjects by chance, rather than by choice, to a treatment or control group. For example, if a psychologist is trying to determine if a violent TV shows cause aggressive behavior in adolescents, she may collect a group of volunteer subjects and randomly divide them into control and treatment groups. The treatment group may watch violent TV shows, and the control group may watch calm TV shows. By randomly dividing the test subjects into these two groups, and not allowing them to choose which group to participate in, we can collect unbiased results from this experiment. If we instead let the test subjects choose to watch either violent or calm TV shows, we would not get an accurate collection of data.

Sometimes, natural experiments can occur in everyday settings. You don't always need a formal experimental setting to observe the results of an experiment. A natural experiment is an empirical study in which some process randomly assigns subjects to control and treatment groups. For example, Melissa Dell, and economist who just recently won the bates Clark prize, has explored the natural experiment to determine the effect of different bombing policies during the Vietnam war. Sometimes, random factors caused some villages to be bombed and others are spared. Those that are spared tend to have different economic outcomes. Melissa Dell studied these outcomes. Obviously, this kind of natural experiment would never be approved to be performed intentionally by a research group. Aside from the fact that it is completely unethical to conduct this kind of research on test subjects, the loss of human life would be so great that it would outweigh the research findings of the study.

2 Week 2

2.1 Optimization

Optimization Using Total Value

Optimization is the first principle of economics. Economists use optimization to predict the choices of that people, households, businesses, and governments. Optimization is not easy, it is quite complex. For example, trying to choose an apartment in a city is a difficult decision. There are multiple factors included in the decision, such as cost, commuting time, square footage, views, etc. We will learn two techniques of optimization in this chapter. Both methods will yield the same results. However, it's important to learn both.

1. Optimization using total value.

   This method calculate the total value of each feasible option and then picks the option with the greatest total value.

2. Optimization using marginal analysis.

   This method focuses on the differences among the feasible options and finds the best option by analyzing these differences.

Let's apply the principle of optimization using total value to the decision of choosing an apartment in a city.

Optimization using total value includes the following three steps:

1. Translate all costs and benefits into common units, like dollars per month.
2. Calculate the total net benefit of each alternative.
3. Pick the alternative with the highest net benefit.

Let's say that we have narrowed our options down to a list of four apartments. These are all located within the same city, but are different distances from the central business district. This means that the commuting time from each apartment differs. Each apartment also differs in its monthly cost of rent. Our short list is below.

Apartment	Commuting Time (hours/month)	Rent ($/month)
Very Close	5 hours	$1,180
Close	10 hours	$1,090
Far	15 hours	$1,030
Very Far	20 hours	$1,000

Figure 2.1: Apartment Options

These apartments are all the same in terms of size and views (one bedroom apartments with a city view.) We are going to evaluate the alternative with the lowest cost. First, we need to

find the total cost of the apartment. This includes the direct and indirect costs. This includes the monthly cost of rent (direct) and the monthly cost of commuting time (indirect). We can compute the monthly cost of commuting time by computing the total wages we could have earned in the time that we spent commuting from our apartment. For example, if we make $10 an hour, this is the hourly value of the alternative activity that we could've been doing instead of commuting. The monthly costs are computed below.

Apartment	Commuting Time (hours/month)	Commuting Cost ($/month)	Rent ($/month)	Total Cost: Rent + Commuting ($/month)
Very Close	5 hours	$50	$1,180	$1,230
Close	10 hours	$100	$1,090	$1,190
Far	**15 hours**	**$150**	**$1,030**	**$1,180**
Very Far	20 hours	$200	$1,000	$1,200

Figure 2.2: Apartment Options and Total Costs

From the table above, we can see that the "far" apartment has the lowest cost out of all four options. This cost is taking into account both rent and commuting time. We can graph this below to visually see how this apartment compares to other apartments.

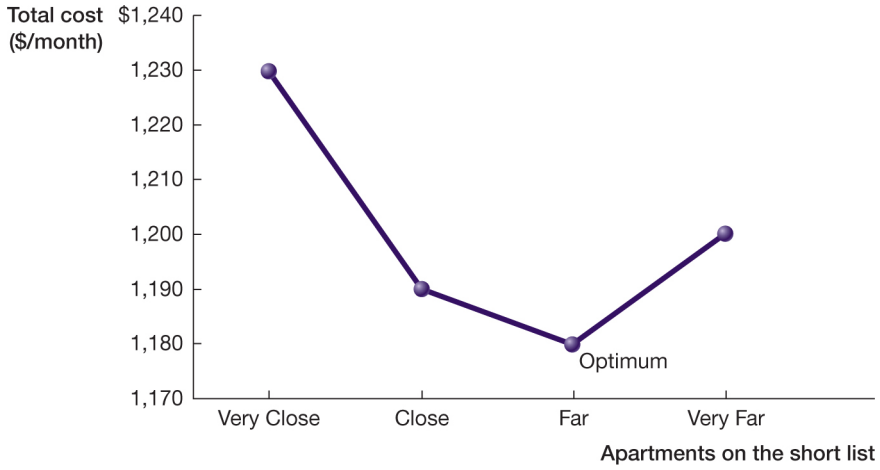

Figure 2.3: Graphing Total Costs

Now what if your hourly wage increased to $15 an hour? This would potentially change our optimal choice. The optimal choice is also referred to as the optimum by economists. We can then compute the commuting cost with this new wage of $15 per hour.

Apartment	Commuting Time (hours/month)	Commuting Cost ($/month)	Rent ($/month)	Total Cost: Rent + Commuting ($/month)
Very Close	5 hours	$75	$1,180	$1,255
Close	10 hours	$150	$1,090	$1,240
Far	15 hours	$225	$1,030	$1,255
Very Far	20 hours	$300	$1,000	$1,300

Figure 2.4: Apartments and Total Costs for A $15 Wage

Given these new calculations, the optimum is now the "close" apartment. This is because this option offers the lowest costs.

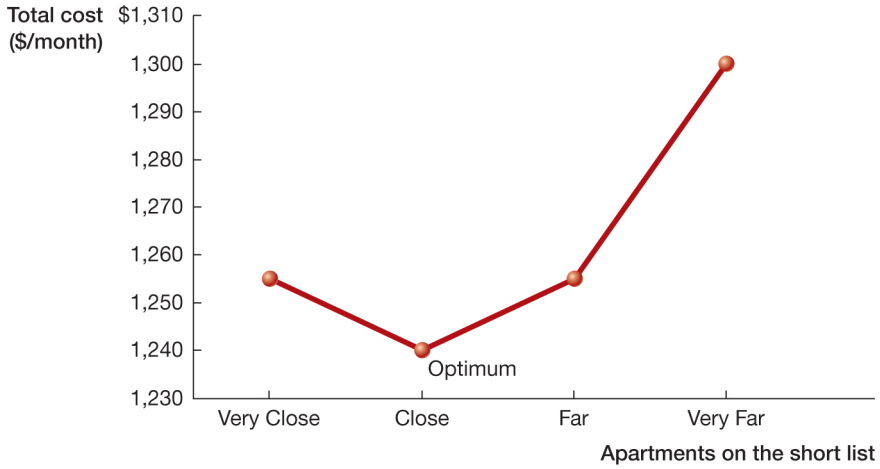

Figure 2.5: Graphing Apartments and Total Costs for A $15 Wage

Optimization Using Marginal Analysis

Optimization using marginal analysis includes three steps:

1. Translate all costs and benefits into common units, like dollars per month.
2. Calculate the marginal costs of moving between alternatives.
3. Choose the option in which the lowest marginal total cost.

Optimization using marginal analysis is often faster to implement compared to the total value method because the marginal analysis method focuses only on the differences between alternatives. The marginal analysis method thinks about how costs and benefits change as you hypothetically move from one alternative to another. For example, let's say you're considering taking a four day or five day vacation. In applying optimization using marginal analysis, you would think about the differences between the two trips. An optimizer would choose the five day vacation if the benefit of vacationing for the fifth day exceeds the cost of vacationing for the fifth day. In choosing between the two options, the optimizer does not necessarily worry about the first four days. The optimizer focuses on one thing that differentiates between the two options, and that is the fifth day.

Economists use the word marginal to indicate a difference between alternatives, usually a difference that represents one step or unit more. The fifth day of vacation is the difference, or margin, between the two trips. Marginal analysis compares the costs and benefits of doing one step more.

In applying this method to the apartment decision, we can evaluate the change in costs and benefits as we move further away from the central business district. Let's assume we have a $10 hourly wage. We are looking at the same shortlist of four apartment options that all differ in monthly rents and commuting times. We can compute the marginal costs in the three columns on the right side of the table below.

Apartment	Commuting Cost	Marginal Commuting Cost	Rental Cost	Marginal Rental Cost	Total Cost	Marginal Total Cost
Very Close	$50	—	$1,180	—	$1,230	—
Close	$100	$50	$1,090	−$90	$1,190	−$40
Far	**$150**	**$50**	**$1,030**	**−$60**	**$1,180**	**−$10**
Very Far	$200	$50	$1,000	−$30	$1,200	$20

Figure 2.6: Marginal Costs Assuming a $10 Opportunity Cost

The marginal commuting cost between the "close" and "very close" apartments is $50 because the difference in commuting costs is $50 ($100 -$50). The marginal rental cost between the close and very close apartments is -$90 because the difference in rents is -$90 ($1090-$1180). We can then compute the marginal total cost in the last column on the right. The marginal total cost between a close and very close apartments is -$40 because the cost drops from $1190 - $1230. The marginal cost is the extra cost generated by moving from one feasible alternative to the next feasible alternative. We can interpret the last column as such: the total cost falls by $40 when we move from the very close to the close apartment.

The cost minimizing choice in this case is the "far" apartment. This is the optimum. This is because moving from close to far made the worker better off. However, moving from far to very far made the worker worse off. Far is the only apartment that satisfies the following property: moving to the apartment makes the worker better off and moving away from the apartment makes the worker worse off. Essentially, we want to choose the option that is just on the brink of the marginal cost turning positive. In this case, the very far apartment has a positive marginal cost. This means that the very far apartment has a higher cost than the other options. However, the far option is the last alternative that provides a drop in marginal total costs. We can see this in the graph below.

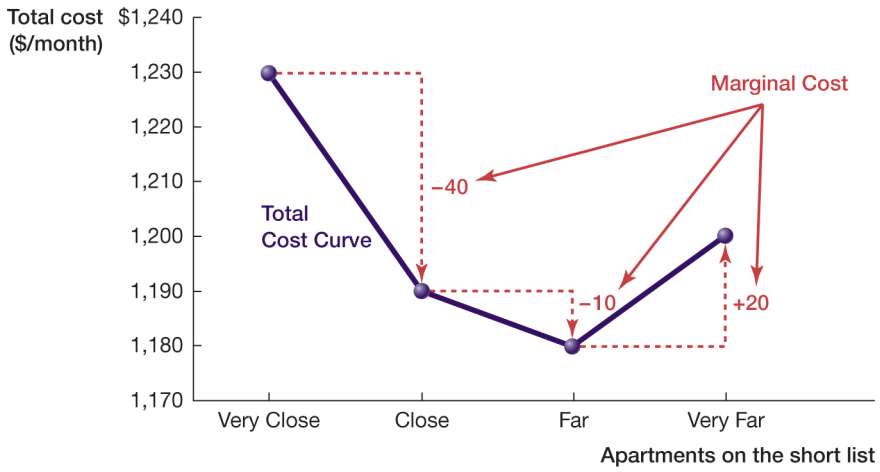

Figure 2.7: Graph of Marginal Cost of Apartments

Marginal analysis means that the optimal choice is the last point in which the the marginal cost is negative. This means that we should keep going down the list of alternatives until we reach the point in which the marginal cost is negative, right before it turns positive. Where the total cost is falling, marginal cost will be negative. Where the total cost is increasing, marginal cost will be positive. We choose the last point at which marginal cost is still negative if we are reading the total cost curve from left the right. Optimization using a marginal analysis will always pick out a single optimal alternative when the total cost curve has the bowl like shape in the graph above.

Although both methods of optimization reach the same conclusion, most economists use marginal analysis in practice. Optimization at the margin is simple because you can ignore everything about alternatives that are being compared except for the particular attributes that are different. Marginal analysis reminds you to exclude information that is irrelevant to your decision.

2.2 Demand, Supply and Equilibrium

Markets

A market is a group of economic agents or trading a good or service plus the rules and arrangements for trading. An example of a market is the market for gas. Every year, over 1 billion drivers visit gas stations globally. Every gas station has the capacity for each person to fill their tank. You'll almost never find an empty gas station that does not offer gas. A market may have a specific physical location, but it is not required to have one to be a market. Although in the gas station example, there is a specific physical location to acquire gas, that is not the case in every market. For example, you can purchase goods and services using Wi-Fi. You do not need to be in the physical location of the item you are purchasing, the transaction can easily be done online. In this chapter, we will focus on markets in which exchanges occur voluntarily at flexible prices determined by market forces. Prices are important because they encourage trade between sellers who can produce goods at relatively low costs and buyers who place a relatively high value on goods.

In this chapter, we will focus on a specific type of market. This is a perfectly competitive market. A **perfectly competitive market** is a market in which all sellers produce an identical good or service and any individual buyer or individual seller is not powerful enough to dominate the market price of that good or service. If all sellers and buyers face the same price, this is the market price. Although there are several different types of markets, we will focus on a perfectly competitive market. A good real life example of a perfectly competitive market is the gas market. Each gas station produces the same type of fuel, and none of them are powerful enough to dominate the field price. Therefore, we see the same cost of fuel at every gas station. Although there are slight differences between certain locations, there is no single gas station that can dominate the market. I think the only gas station that could come close to dominating the market with low prices would be Costco fuel. However, this is an exclusive market for customers with Costco memberships. Another important element of a perfectly competitive market is that buyers and sellers are all price takers. A **price taker** is a buyer or seller who accepts the market price. In other words, buyers are unable to bargain for a lower price and sellers cannot bargain for a higher price in a perfectly competitive market.

Buyer Behavior

Prices can change buyer behavior. For example, higher gas prices may deter drivers from filling their tanks and instead prefer to ride her bike to work. **Quantity demanded** is the amount of a good that buyers are willing to purchase at a given price. A **demand schedule** is a table that reports the quantity demanded at different prices, assuming all other variables are equal. There is a negative relationship between price and quantity demanded. In the gasoline example, the quantity of gasoline purchased decreases as price increases. We can see this in the graph below. As the price per gallon reaches seven dollars, the number of gallons purchased each year decreases. This is very applicable to current rising gas prices that we see here in Reno.

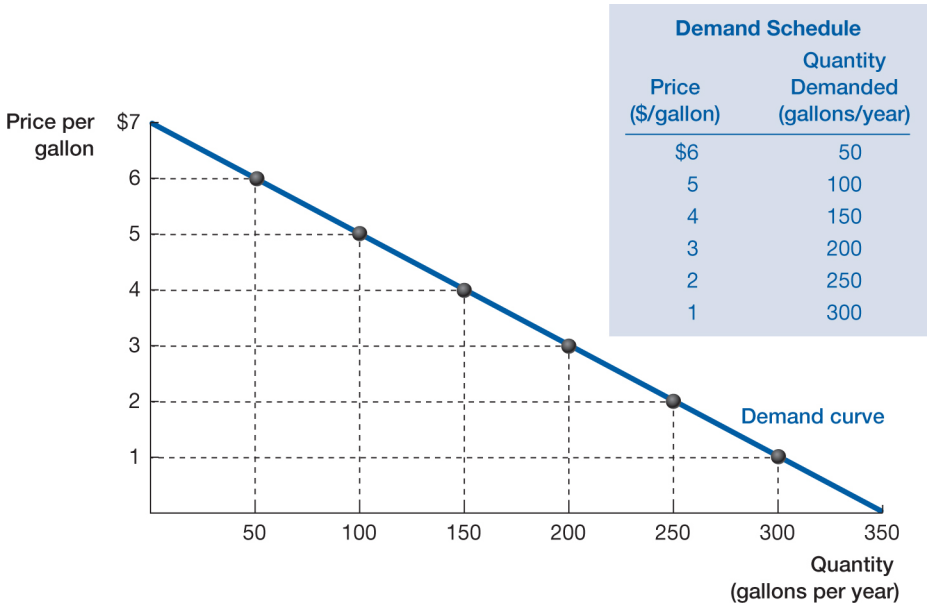

Figure 2.8: Demand Schedule for Gasoline

The demand curve plots the relationship between the quantity demanded and the price of a good, which is the demand schedule. Each point on the demand curve represents the quantity of gallons purchased in a year at that price. For example, when gas is seven dollars a gallon, this individual will not purchase any gas. When gas is six dollars a gallon, this individual will purchase 50 gallons. Each of us has a unique demand curve.

The negative relationship between price and quantity demanded is an example of the law of demand. The law of demand states that the quantity demanded rises when the price falls, holding all else equal. "Holding all else equal" is a fancy phrase to establish that we are assuming the product of the quality is the same, the environment of the market is the same, and nothing significant has changed in the world. In other words, the only thing that's changing is price. Everything else is equal to the same thing as before, in other words we are holding all else equal!

Another piece of information represented in the demand curve is a consumer's willingness to pay. Willingness to pay is the highest price that a buyer is willing to pay for an extra unit of a good. For example, according to the demand curve above, if that individual has already purchased 149 gallons of gas that year, then she is willing to pay four dollars for the 150th unit of gas. If she then purchased 199 gallons of gas that year, She is willing to pay three dollars for that 200th gallon of gas. The downward sloping demand curve demonstrates the idea of a diminishing marginal benefit. Diminishing marginal benefit means that as you consume more of a good, your willingness to pay for an additional unit of that good declines. You can understand this idea while eating an entire pizza. Your first slice of pizza makes you incredibly happy. Your second slice of pizza still makes you happy, but not as happy as the first. By the time you get to your fifth slice of pizza, you were so full that it does not provide you as much happiness as your first slice of pizza. This is a diminishing marginal benefit.

The demand curve in the graph above illustrates the demand schedule for a single individual. However, in the real world, millions of individuals demand gasoline. We can represent this aggregated demand in a demand curve. By summing up the quantity of gas demanded at various price levels, we can illustrate the market demand curve. The market demand curve is the sum of the individual demand curves of all potential buyers. It plots the relationship between the total quantity demanded and the market price, holding all else equal. We can see this in the graph below. Let's take two individual consumers, such as Sue and Carlos. Let's assume that these two individuals are the only consumers in the market for gas. Obviously, there are more than two people in the gas market in the real world. However, just for simplicity purposes, we'll assume that there are only two in this market.

Sue's Demand Schedule

Price ($/gallon)	Quantity Demanded (gallons/year)
$5	100
4	200
3	300
2	400
1	500

Carlos's Demand Schedule

Price ($/gallon)	Quantity Demanded (gallons/year)
$5	200
4	400
3	600
2	800
1	1,000

Total Demand Schedule

Price ($/gallon)	Quantity Demanded (gallons/year)
$5	300
4	600
3	900
2	1,200
1	1,500

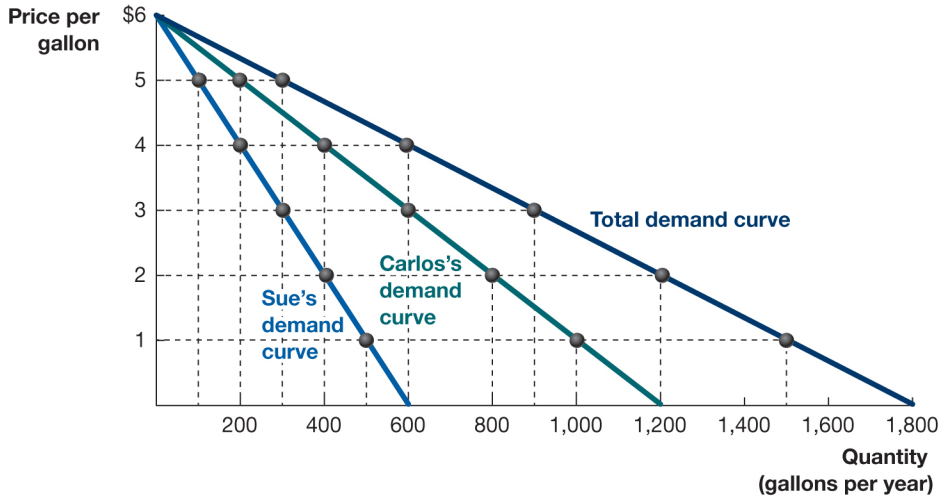

Figure 2.9: Aggregation of Demand Schedules and Demand Curves

According to the demand schedules above, at $5 a gallon, Sue only buys 100 gallons of gas. Carlos buys 200 gallons of gas. Therefore, the aggregate quantity of gas demanded is 300 gallons at $5 a gallon. We can do the same computations for each price level. This helps us to construct an entire demand curve.

It is possible for the demand curve to shift positions. It can shift to the right or to the left. It will always maintain its downward sloping shape, but it can shift its' position on the axes. A shift in the demand curve is caused by anything other than a change in prices. It's important to remember that a change in prices prompts a movement along the demand curve. For example, at $5 a gallon,

Sue will purchase 100 gallons of gas. At $4 a gallon, Sue will purchase 200 gallons of gas. This is a movement along the demand curve. However, we will now focus on how to shift the demand curve entirely.

There are several things that can shift the demand curve:

1. Tastes and preferences

   Consumer preferences evolve over time. For example, during the beginning of the COVID-19 pandemic in March 2020, there was a significant decrease in number of commuters on the road. This is because many people worked from home during the lockdowns. Therefore, there was a shift in demand for gas, regardless of prices. This kind of drop in demand would prompt a leftward shift of the demand curve. We can see this in the graph below. Given the change in consumer preferences, This prompted a leftward shift of the demand curve.

2. Income and wealth

   A change in an individual income can affect their quantity demanded of certain goods. For example, an increase in income might cause me to spend more money on dining out instead of making home cooked meals. However, a decrease in income might cause me to spend more money on home-cooked meals instead of dining out. This is a great example of normal versus inferior goods. When income increases, individuals demand more of a normal good. The normal good in this case is dining out. When I have a higher income, I like to spend more money on dining out for dinner. The inferior goods in this case include home-cooked meals. When my income decreases, I demand more of these home-cooked meals because they are cheaper. An increase in income shifts the demand curve to the right for a normal good and to the left for an inferior good. A decrease in income shifts the demand curve to the left for a normal good and to the right for an inferior good.

3. Availability and prices of related goods

   Demand for a good can shift based on available, cheaper alternatives. For example, demand for gas may decrease if public transportation is available and cheap. Two goods are substitutes when a rise in the price of one good leads to a rightward shift in the demand curve for the other. Another example of substitutes includes Nike and Adidas shoes. If I typically purchase Nikes, I may change my mind if Adidas significantly drops their prices. Adidas shoes are a cheaper alternative to Nikes. Two goods can also be complements if the fall in the price of one good leads to a rightward shift in demand for the other. For example, if hot dogs become cheaper, this may influence people to purchase more hot dogs, and consequently purchase more ketchup. Ketchup and hot dogs are complements. They go well together. A decrease in the price of hotdogs may lead to an increase in the demand for ketchup. Therefore, the availability of hot dogs can shift the demand curve for ketchup.

4. Number and scale of buyers

   The number of buyers in a market can shift the demand curve. When the number of buyers increases, the demand curve shifts to the right. When the number of buyers decreases, the demand curve shifts to the left. For example, the demand curve for air purifiers significantly increases during fire season for Reno. The number of consumers purchasing an air purifier

in the month of August and September increases dramatically. The demand curve for air purifiers shifts to the right in this case because there are more buyers, and therefore a higher quantity of air purifiers demanded.

5. Buyers' beliefs about the future

   Changes in buyers' expectations about the future can shift the demand curve. For example, if consumers lose their jobs during an economic recession, they may decrease the demand for weekend trips to local ski resorts.

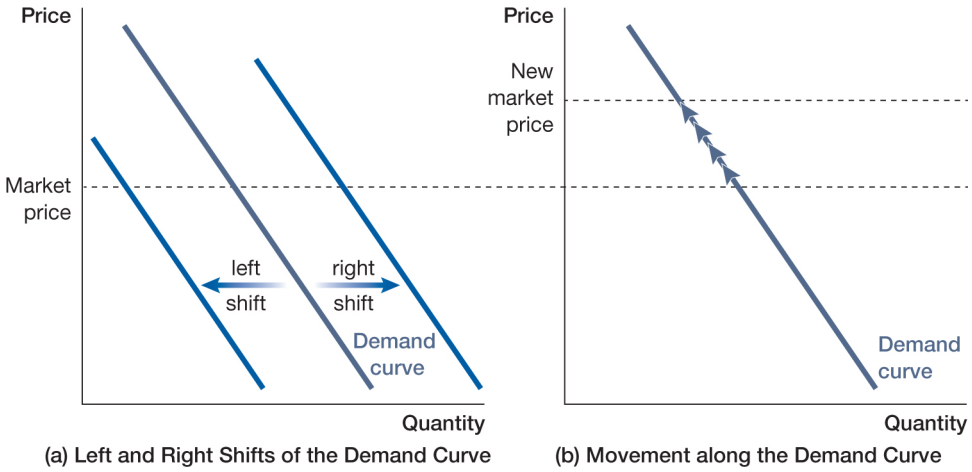

Figure 2.10: Shifts vs. Movements Along the Demand Curve

Seller Behavior

Now that we understand buyer behavior, we need to understand seller behavior in order to analyze markets. Quantity supplied is the amount of a good or service that sellers are willing to sell at a given price. For example, not all sellers are willing to engage in a market if they cannot sell their goods for a certain price. For example, Exxon mobile is not willing to drill for oil unless they can sell a barrel of oil for over $60. It takes an intense amount of work to drill for oil, so they must sell each barrel for at least $60 in order to pay off all of their expenses.

Just like with a demand schedule, we can construct a supply schedule. A supply schedule is a table that reports the quantity supplied at different prices, assuming everything else is equal. The supply curve plots the quantity supplied at different prices. A supply curve plots the supply schedule. We can see this in the graph below. As the price of the item sold increases, the quantity supplied also increases. These two variables have a positive relationship. This is a demonstration of the law of supply. The law of supply states that as price increases, so does the quantity supplied. If producers can sell their goods at a higher price, they are willing to supply a higher quantity. We can see this in the graph below. As the price per barrel increases, the quantity of barrels supplied also increases. We can see this in the fact that the supply curve is upward sloping.

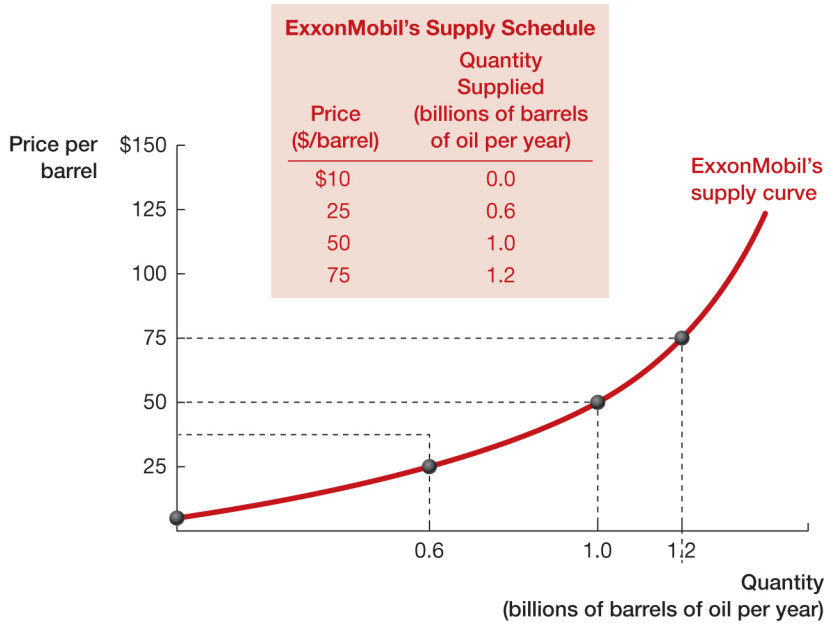

Figure 2.11: ExxonMobil's Supply Schedule

Another piece of information we can learn by looking at the supply schedule is the sellers willingness to accept price. The willingness to accept price is the lowest price that a seller is willing to get paid to someone extra unit of a good. At a particular quantity supplied, willingness to accept is the height of the supply curve. Willingness to accept is also known as the marginal cost of production. For example, according to the Exxon mobile supply curve, Exxon is willing to accept $50 to produce its 1 billionth barrel of oil. It is willing to accept $75 to produce its 1.2 billionth barrel of oil.

Just like with aggregate demand, we can also compute aggregate supply. By summing the number of gallons produced at each price level, we can construct the market supply curve. The market supply curve is the sum of the individual supply curves of all potential sellers. It plots the relationship between the total quantity supplied in the market price, holding all else equal. In the graphs below, we can compute the total supply schedule. For example, at $25 per barrel, Chevron will produce 0.4 billion barrels of oil per year. ExxonMobil will produce 0.6 billion barrels of oil per year. The sum of these two numbers represents the aggregate quantity supplied, which is 1 billion barrels of oil per year at the price of $25 per barrel. We can then plot the total supply curve.

Chevron's Supply Schedule

Price ($/barrel)	Quantity Supplied (billions of barrels of oil per year)
$10	0.0
25	0.4
50	0.7
75	0.9

ExxonMobil's Supply Schedule

Price ($/barrel)	Quantity Supplied (billions of barrels of oil per year)
$10	0.0
25	0.6
50	1.0
75	1.2

Total Supply Schedule

Price ($/barrel)	Quantity Supplied (billions of barrels of oil per year)
$10	0.0
25	1.0
50	1.7
75	2.1

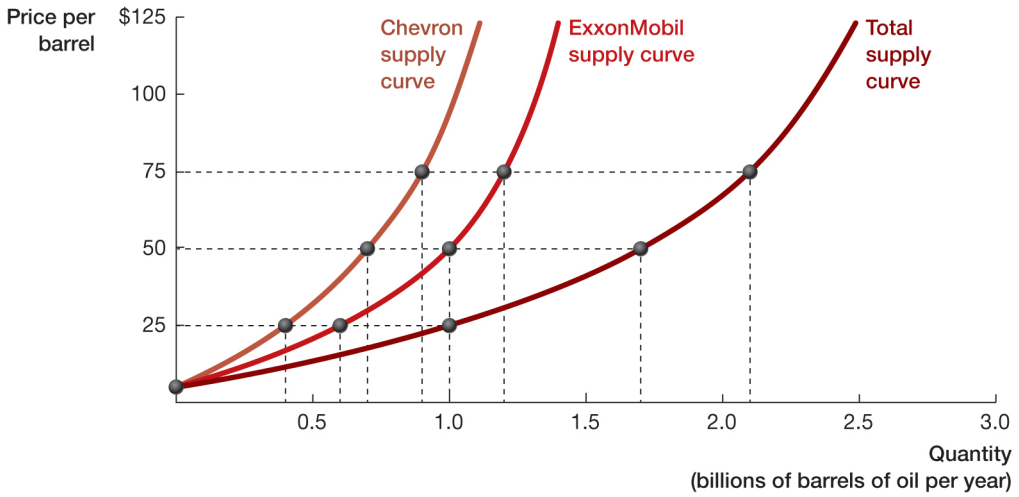

Figure 2.12: Aggregation of Supply Schedules

Just like how demand can shift, supply can also shift. Remember, a change in the selling price of the final good or service prompts movements along the supply curve whereas a change in outside variables prompts a shift in the supply curve. There are several factors that can shift the supply curve. These factors all sound very similar to the factors that can shift the demand curve.

1. Prices of inputs used to produce the good

   An input is a good or service used to produce another good or service. Changes in the prices of inputs can shift the supply curve. For example, steel is a critical input in the production of oil. An increase in the price of steel implies that some opportunities to produce oil will no longer be profitable. Therefore, oil producers will choose to supply a reduced amount of oil. Therefore, at $25 a barrel, instead of producing 0.4 billions of barrels of oil per year, Chevron may choose to produce 0.2 billions of barrels oil per year if the cost of steel increases dramatically. This would prompt a leftward shift of the supply curve.

2. Technology used to produce the good

   Technology can improve the productivity of producers. For example, fracking has revolutionized the energy industry. This technology increases the productivity of oil producers, prompting a sharp rightward shift of the supply curve for petroleum and natural gas.

3. Number and scale of sellers

The number of sellers in a market can shift the supply curve. For example, a dramatic drop in the number of oil producers can dramatically shift the supply curve to the left. A real life example of this was in 2011 when Libyan citizens overthrew their dictator. Fighting dragged on for months after this. During this period, Libya essentially stopped oil production and the supply curve shifted to the left.

4. Sellers' beliefs about the future

   If sellers expect demand to increase in the future, they may preserve stock piles of production in order to prepare for this. For example, natural gas usage increases dramatically every winter. This creates a winter spike in natural gas prices. Natural gas producers expect this increase in demand in prepare by storing a vast quantities of gas during the summer months. This implies that the supply of natural gas shifts to the left during the summer months. By pulling supply from the summer market and storing it to sell later in the winter market, natural gas suppliers obtain a higher average price over the full year.

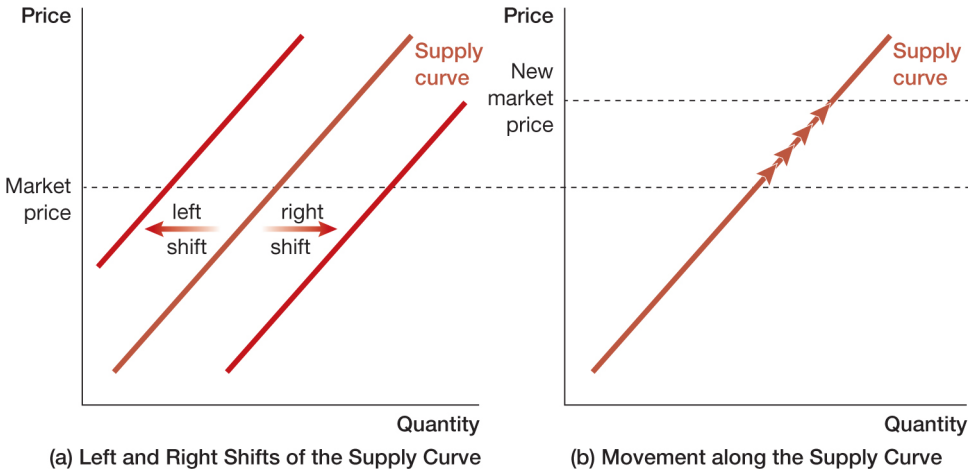

Figure 2.13: Shifts vs. Movements along the Supply Curve

Supply and Demand in Equilibrium

Buyers and sellers engage in a market. Competitive markets converge to the price at which quantity supplied equals quantity demanded. To visualize this understanding, we need to graph the supply and demand for together on the same graph. Price is on the Y axis and quantity is on the X axis. We can plot the aggregate supply curve in the aggregate demand curve for oil. The X axis represents the quantity in billions of barrels of oil per year. The Y axis represents the price per barrel. The demand curve is in blue and the supply curve is in red. The two curves cross at a price of $50 per barrel and a quantity of 35 billion barrels. This crossing point is the competitive equilibrium. At this competitive equilibrium price, the quantity supplied equals quantity demanded. This is sometimes referred to as the market clearing price. This is the price in which there is a buyer for every unit that is supplied in the market. The quantity here is referred to as a competitive equilibrium quantity.

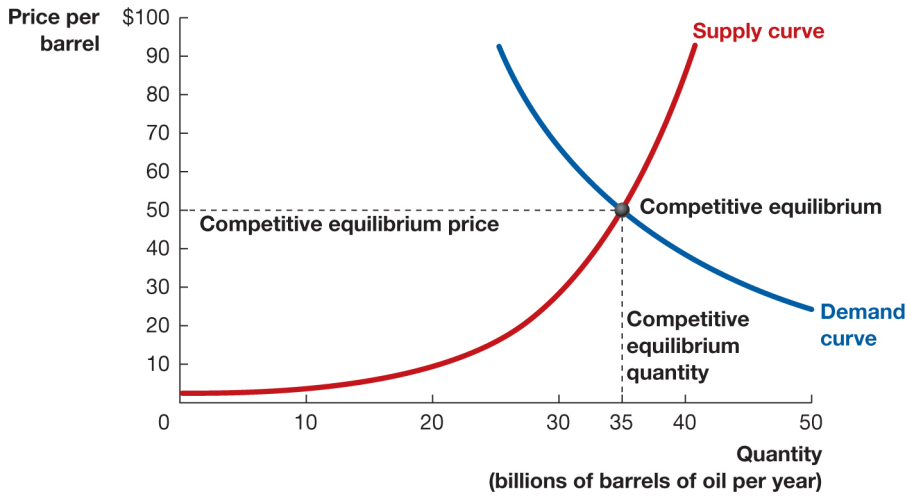

Figure 2.14: Demand and Supply for Oil

If the market price is above the competitive equilibrium price, then the market is not in equilibrium. When the market price is above the competitive equilibrium price, the quantity supplied exceeds the quantity demanded, creating excess supply. We can see an example of this in the graph below. If the market price is at $70, but the competitive equilibrium price is at $50, then we will encounter a case of excess supply. In this case, the quantity supplied exceeds the quantity demanded. Therefore, the oil producers will have leftover oil barrels on their shelves because no one wants to buy them. However, this dynamic would naturally push down the market price of oil until it hits the competitive equilibrium price of $50 per barrel. This is because existing oil storage tanks are limited and expensive to build. Therefore, sellers would begin discounting their barrels of oil in order to get rid of the excess supply.

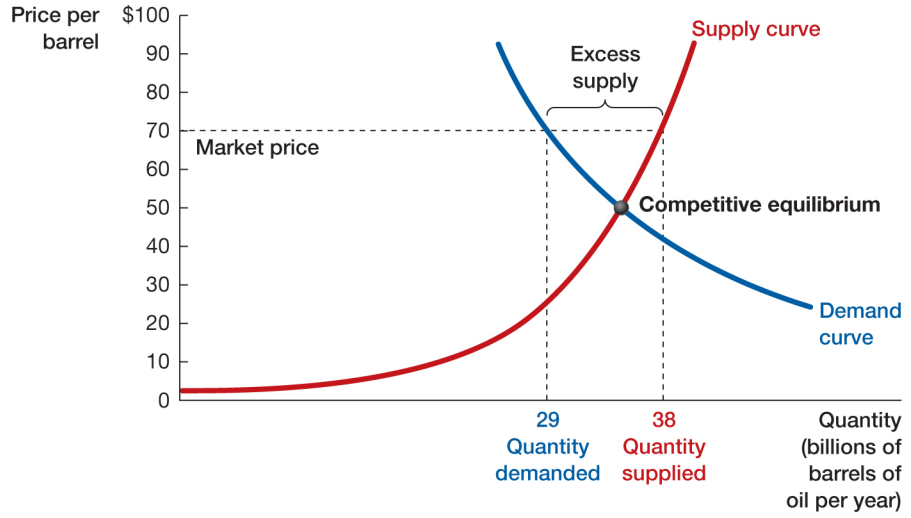

Figure 2.15: Excess Supply

It is also possible to encounter a case of excess demand in which the market price is below the competitive equilibrium price. In this case, the quantity demanded exceeds the quantity supplied.

Therefore, producers cannot produce enough supply and there is a shortage on the shelves. We can see in the graph below, in the case of this excess demand, there are 44 billion barrels of oil demanded per year but there are only 30 billion barrels produced each year.

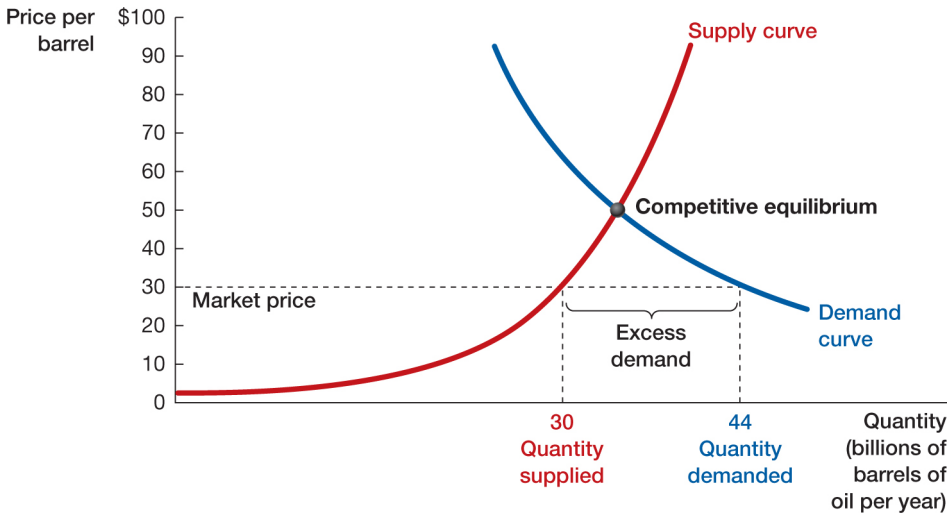

Figure 2.16: Excess Demand

So what happens to the market equilibrium when there is a shift in either the supply or demand curves? We can plot this shift in the graphs below, and see the consequential change in market equilibrium price.

For example, what would happen if a major oil exporter suddenly stopped producing, such as mini producers did in Libya in 2011? This would cost a shift to the left of the supply curve, and a drop in the supply of oil. Consequently, the price of oil needs to rise from its old level so that quantity supplied can equal quantity demanded. The rise in the equilibrium price is an example of a movement along the demand curve. Because the demand curve has a downward slope, a rise in price causes a reduction in the quantity demanded.

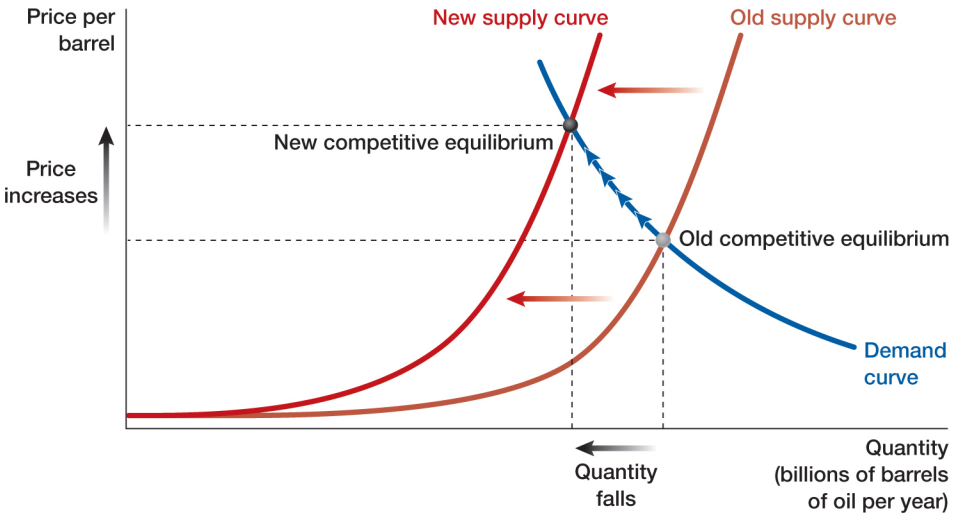

Figure 2.17: Leftward Shift of Supply Curve

Now what if the supply curve instead shifted to the right? Let's say there was a technological breakthrough and producers were suddenly able to supply more oil to the market. This would mean that they would be an increase in the supply of oil and consequently a shift to the right of the supply curve. Therefore, the price of oil needs to fall in order for the quantity supplied to equal the quantity demanded. This fall in the equilibrium oil price is an example of a movement along the demand curve. Because the demand curve has a downward slope, a drop in equilibrium price causes an increase in the quantity demanded.

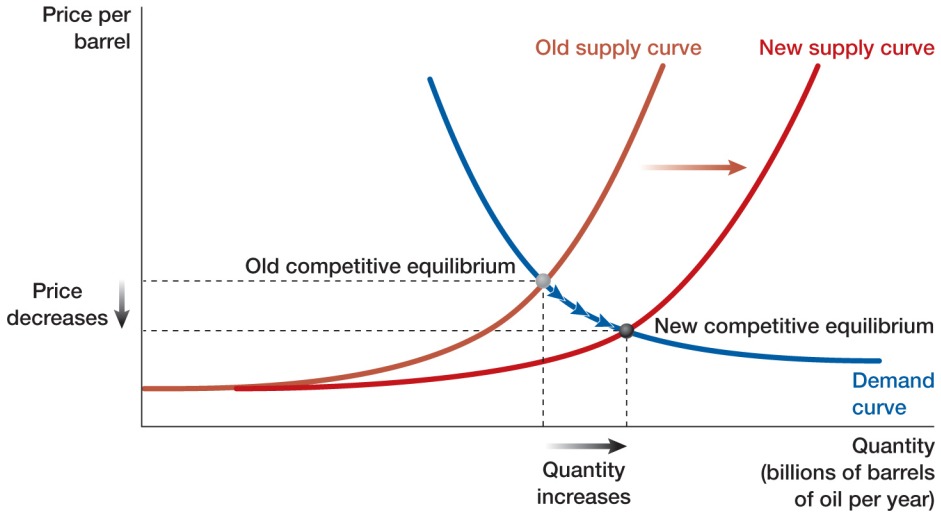

Figure 2.18: Rightward Shift of Supply Curve

Now what if there is a shift in the demand curve? Let's say overall consumer preferences for crude oil have declined and this causes a leftward shift of the demand curve. Consequently, the equilibrium price of oil needs to fall from its old level in order for quantity supplied to equal quantity demanded. Because the supply curve is upward sloping, a decline in price causes a reduction in the quantity supplied. You'll notice that a fall in price also causes a fall in the equilibrium quantity.

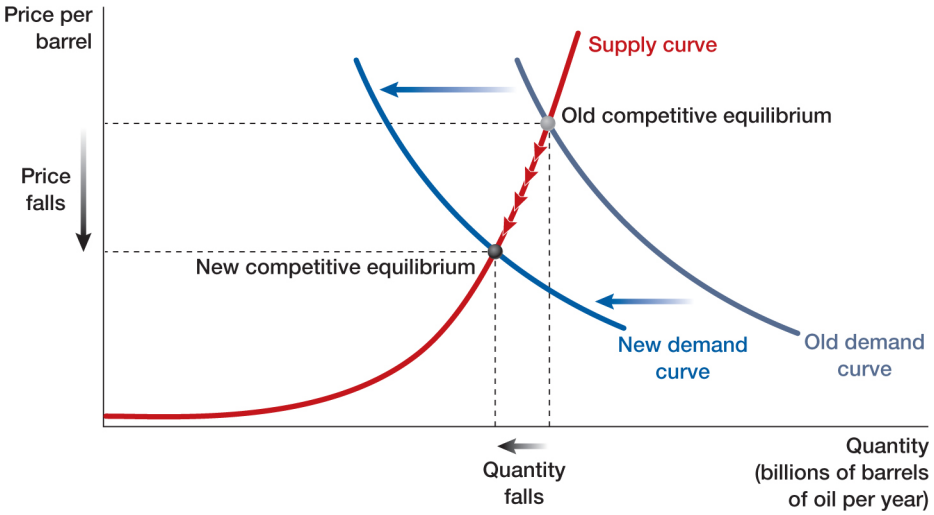

Figure 2.19: Leftward Shift of Demand Curve

It is also a possibility for both the demand and supply curves to shift simultaneously. For example, the fracking revolution has shifted the supply curve for oil to the right. At the same time, the increase in environmental consciousness and energy saving technology has shifted the demand curve for oil to the left. We can see this in action in the graph below. When the demand curve shifts to the left and the supply curve shifts to the right, the competitive equilibrium price will always decrease. However, the competitive equilibrium quantity may decrease, stay the same, or increase overall. The old demand curve is labeled D1 and the new demand curve is labeled D2. The old supply curve is S1 and the new supply curve is S2. The gray dot represents the old equilibrium and the black dot represents the new equilibrium. In all three panels, the new equilibrium price is lower than the original equilibrium price. However, the new equilibrium quantity varies.

In the panel a, the leftward shift in the demand curve dominates in the equilibrium quantity falls from Q1 to Q2. In panel B, the equilibrium quantity stays the same period in panel C, the right shift in supply dominates in the equilibrium quantity increases from Q1 to Q2. The key to determining how the equilibrium quantity changes is by comparing the shift in the supply curves to the shift in the demand curves. If the shift in the demand curves is larger than a shift in the supply curves, the equilibrium quantity will decrease. If the shift in the demand curves is just about the same as the shift in the supply curves, then the equilibrium quantity will remain the same. If the shift in the demand curves is smaller then the shift in the supply curves, then the equilibrium quantity will increase.

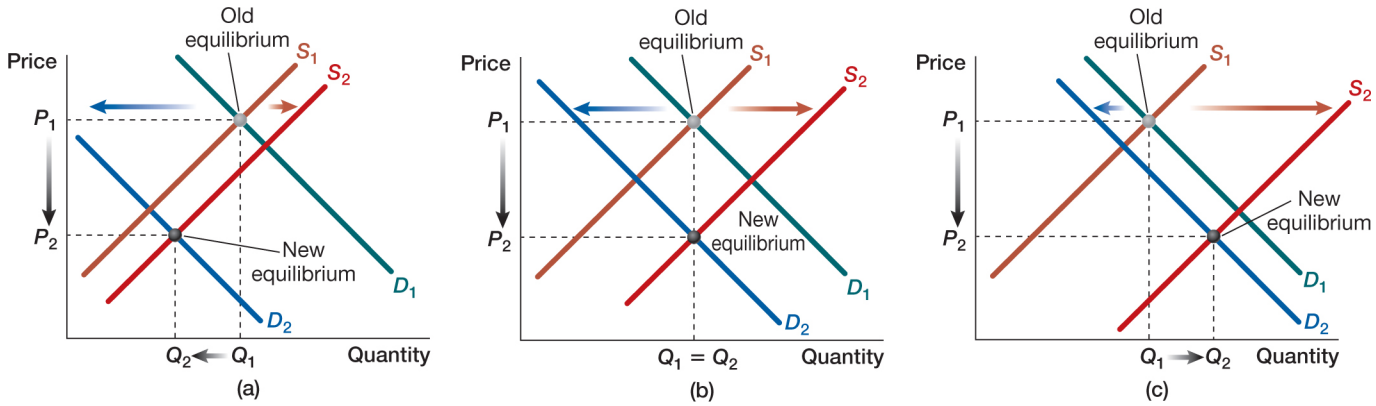

Figure 2.20: Demand Curve Shifts Left and Supply Curve Shifts Right

3 Week 3

3.1 Macroeconomic Aggregate Metrics

Macroeconomic Questions

Macroeconomics is the study of economic aggregates and economy wide phenomena, the annual growth rate of a country's total economic output, or the annual percentage increase in the overall cost of living. Macroeconomists are interested in the differences in income across countries and policies that would enable countries to sustainably increase their level of economic output.

One thing macroeconomists like to study is income per capita. Income per capita is income per person. It is calculated by dividing a nation's aggregate income by the number of people in the country. In other words, income per capita represents the average income per person. The income per capita in the United States is almost 4 times the income per capita in China and almost 100 times the income per capita in Zimbabwe.

Some questions that are of interest to most macroeconomists include: will china eventually match the level of US income per capita? Will it surpass the United States? What can be done to improve the living conditions in impoverished nations like Zimbabwe? Annual income per capita in Zimbabwe was $1900 in 2018, which means that families are struggling to maintain basic levels of food and housing needs.

There are some important terms to know before we move forward:

1. An economic downturn that lasts for at least two quarters is a **recession.**

2. During recessions, the unemployment rate increases. A person is officially **unemployed** if three conditions are satisfied: 1) he or she does not have a job, 2) he or she has actively looked for work in the prior four weeks, and 3) he or she is currently available for work.

3. The **unemployment rate** is the fraction of the labor force that is unemployed.

4. **National income accounts** measure the level of aggregate economic activity in a country. We'll dive into detail in this pretty soon.

5. **National income and product accounts** includes the system of national income accounts that is used by the government.

We have all experienced economic fluctuations in the United States economy. For example, the economic recession that ran from 2007 to 2009 demonstrated an increase in the unemployment rate. The unemployment rate increased from 4.4% to 10%. The US economy as a whole shrank by 4.3%.

National Income Accounts

There are three main components of the national income accounts that we need to evaluate: production, expenditure and income.

Let's say we live in a country that only produces Ford vehicles. To determine the market value of production in this country, we multiply the quantity of cars produced by the market price of each car. If this country produces 5 million cars each year that are each worth $30,000, this means the country produced a total market value of $150 billion. This is also referred to as the gross domestic product, or the GDP. The gross domestic product is the market value of the final goods and services produced in a country during a given period of time. I want to emphasize that the GDP only includes the value of final goods and services. This means that we count the value of the end product in the chain of production. Components that are used to make these vehicles do not get counted separately because then we would be double counting the value of these products. GDP is a measure of production, not a measure of sales to consumers. Therefore, something that is produced is counted in the GDP, even if it is not sold to a consumer.

Total expenditures include household expenditures on these vehicles in this country. Therefore, if every single car that was manufactured was sold, then total expenditures would equal $150 billion. However, if some vehicles are manufactured but not sold, then those vehicles are considered as purchased by the Ford manufacturer itself.

We can calculate total income by evaluating how much money Ford gives to its workers in the form of wages and how much it gives to its corporate executives who own the company in the form of salaries. For example, we know that Ford generates approximately $150 billion of revenue every year. It gives some of this to their workers and the rest of it to its corporate executives. This means that total income is still $150 billion.

It's not a coincidence that we keep finding this $150 billion value. Every dollar of revenue must either go to a worker or an owner. Therefore, the total value of revenue must equal the total value of income received by workers and owners. This is known as the aggregate accounting identity.

$$\text{production} = \text{expenditure} = \text{income}$$

This is an identity because when two variables are mathematically identical, they create an

identity.

Factors of production are the inputs to the production process. Factors of production come into forms: capital and labor. Both physical capital and labor are owned by households. We need to understand how households and firms interact with one another in order to understand the three parts of the national income accounts: production, expenditure and income. Firms demand physical capital and labor. Firms supply goods and services, like airplanes. Households demand goods and services, like air travel. Households supply physical capital and labor. We can see these interactions in the diagram below.

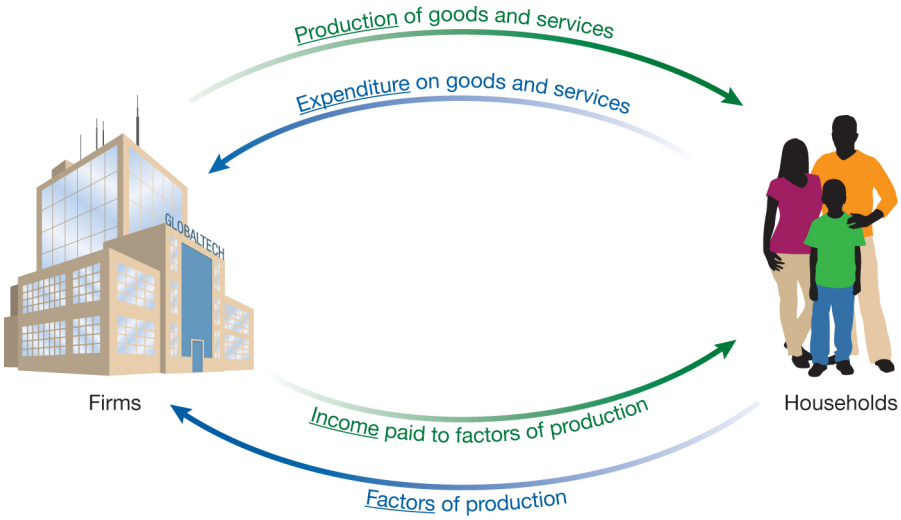

Figure 3.1: Circular Flow Diagram

Production represents the goods and services that are produced by firms. These goods and services are ultimately sold households. Expenditure represents the payment for these goods and services. These payments are made by households to firms. Income represents the payments that are made from firms to households to compensate the household for the use of their physical capital and labor. These payments include things like wages and salaries. Factors of production represents a productive resources that our owned by household and used by firms in the production process.

National Income Accounts: Production

Now we will evaluate production-based national income accounts. Production based accounts sum the market value that is added by each domestic firm in the production process. Production-based accounts measure each firm's value added, which is the firm's sales revenue minus the firm's purchases of intermediate products from other firms. For example, Dell technologies produces computers. Thirty five years ago, Dell assembled almost all of its computers at facilities within the United States. Today, Dell buys most of its computers from foreign manufacturers, especially in Asia. Consider Dell's value added when it sells a laptop for $1000 directly to a US consumer. In this case, Dell pays a foreign supplier $600 in Dallas value added is the difference between the two values, $1000 -$600 = $400. This value added of $400 derives from two factors of production: Dell's domestic employees and Dell's domestic physical capital. Now, let's consider the fact that

Dell sells the same computer at $900 to a third-party retailer like Walmart. Walmart then sells this laptop for $1000. In this case, Dell's value added is $300, which is the difference between $900 and $300. It is important to remember that foreign factories do not count towards U.S. GDP. The production of laptops in a foreign factory is part of a foreign country's GDP because the factory is in a foreign country. Production-based accounting system implies that importing some good from abroad and selling it to a US consumer at the same import price does not add any value. However, importing something for $600 and reselling it for $1000 as a source of production, in the form of $400 in value added. Adding of the value added by all firms in the United States was sum to US GDP.

National Income Accounts: Expenditure

Let's talk about the second mathematically equivalent way to measure GDP, and that is the national income accounts as measured by the purchases of goods and services produced in the domestic economy. There are five categories of these kinds of purchases:

1. **Consumption (C)**

   Consumption is the market value of consumption goods and consumption services that are bought by domestic households. This covers everything from small items like basketballs to large items like vehicles. However, it does not include expenditures that are made on residential construction. It does not include expenditures to buy a pre-existing house or apartment because these types of expenditures are just a transfer of assets from one household to another.

2. **Investment (I)**

   Investment is the market value of new physical capital that is bought by domestic household and firms. New physical capital includes residential houses, business inventories, business structures and business equipment. When macroeconomists talk about investments, they are talking about the purchase of new physical capital, not financial investments such as stocks or bonds.

3. **Government Expenditure (G)**

   Government expenditure is the market value of government purchases of goods and services. This includes tanks, public hospitals, and bridges. For the purposes of the national income accounts, government expenditure excludes transfer payments and also excludes interest paid on government debt. These categories are omitted because they represent payments to other agents in the economy who will use those payments to buy goods and services.

4. **Exports (X)**

   Exports include the market value of all domestically produced goods and services that are purchased by households, firms, and governments in foreign countries. We measure exports in terms of the value added. When a US farmer exports flower to a Japanese supermarket, the value of this export is the price that the US farmer receives from the Japanese supermarket. It is not the price at which the Japanese supermarket sells the flour to Japanese households.

5. **Imports (M)**

   Imports include the market value of all foreign-produced goods and services that are sold to domestic households, domestic firms, and the domestic government. Imports are already counted as part of consumption expenditures, investment expenditures and government expenditures. Therefore, imports overlap with the first three categories in this list. When we calculate US GDP on an expenditure basis, we are going to subtract the value added of imports.

It is important to remember that each purchase appears only once in one of the categories listed in categories 1 through 4.

We're now ready to put together our formal equation to calculate the total value of expenditures on goods and services produced in the domestic economy. This is also known as the national income accounting identity and commonly known as the gross domestic product or the GDP. GDP equation demonstrates at the market value of domestic production is equal to the total expenditure of domestic economic agents, which is the sum of consumption, investment and government expenditure, plus the expenditure of foreign agents on exports from the domestic economy, minus the value of domestic expenditure that was imported.

$$Y = C + I + G + X - M$$

National Income Accounting: Income

Another mathematically equivalent way of measuring the GDP is by income. Income based national accounts track the income of the various agents in the economy. Aggregate income is identical to aggregate production and aggregate expenditure. Therefore, if aggregate expenditure was $21.4 trillion in 2019, and aggregate production and aggregate income were also each $21.4 trillion in 2019.

Income payments come in two categories:

1. **Labor Income**

   Labor income is any form of payment to compensate people for their work. For example, this includes wages, salaries, workers' health insurance and workers' pension benefits.

2. **Capital Income**

   Total income is any form of payment that derives from only physical or financial capital. For example, this includes dividends paid to shareholders, interest paid to lenders, rent payments made to landlords, earnings retained by corporations and the value of your home.

What is not included in the GDP?

Although the GDP is a great way of measuring macro economic activity, it does not take into account several things. It is important to know the components lacking from the GDP.

1. **Physical Capital Depreciation**

   The GDP does not include physical capital depreciation. Depreciation includes the wear and tear a physical capital that causes this capital to lose value over time. For example, driving a truck wears down the brakes and tires over the years. Many governments try to measure depreciation in their national accounts but do not subtract this value from the overall GDP.

2. **Home Production**

   The GDP also does not include home production, which is unpaid work that takes place in the home. For example, if you grow your own flowers, the bouquet that you created is considered a product of home production and is not measured in the GDP. However, if you were to purchase that same bouquet of flowers from a local florist, then it would be included in the GDP. Home production can sum to a significant value, such as all of the unpaid childcare delivered by mothers and fathers in the home.

3. **The Underground Economy**

   There is a whole world of transactions that take place under the radar. For example, paying your plumber or your taxi driver in cash does not show up in the GDP accounts. Although plumbers and taxi drivers are legal occupations, some workers hide income to avoid paying taxes. In the United States, there is an estimate of $500 billion per year in transactions that are not taxed.

4. **Externalities**

   Externalities occur when an economic activity has spillover effects that do not affect those directly engage in the activity. For example, factories that produce excessive air pollution indirectly affect the nearby residential areas that then have to endure this polluted air. Externalities are not included in the GDP.

It is important to remember that the US GDP only includes gross domestic production. Therefore, we are only including the goods and services that are produced within the United States. The **gross national product** is the market value of production generated by the factors of production, both capital and labor, possessed or owned by the residence of a particular nation. For example, the US GNP includes the production of a worker who normally resides in the United States, even if the production occurred when the worker was temporarily working abroad. For example, if a US professor delivered a summer course at the national University of Singapore, her salary, which was paid by the national University of Singapore, would still be included in the US GNP. It would be excluded from Singapore's GNP. US GNP would exclude the value added of machines owned by a Japanese car manufacturer, even if those machines are operated in Alabama. The US GNP is carefully constructed to only include the value added of factors of production possess are owned by US residents, no matter where those factors of productions operate in the world. The formal equation is below:

Gross national product = Gross domestic product + production of US on capital and labor in foreign countries - production of foreign owned capital and labor in the United States

Real vs. Nominal

In order to distinguish between the growth in GDP and the growth in overall prices, we need to learn about real GDP versus nominal GDP. For example, going back to the example country that only produces Ford motor vehicles, if that country produce 10 cars in 2019 and each car was worth $30,000, the total GDP in 2019 would be $300,000. If they produce the same amount of 10 cars in 2020, but the price of each vehicle has increased to $400,000 then the GDP in 2020 is $400,000. Although we can brag that GDP has increased by 33%, It's not an accurate reflection of our domestic production. We did not produce more cars, the only thing that changed was the price of these vehicles. Therefore, we should distinguish between the nominal value of 400,000 and the real value of 300,000. Nominal GDP is the total value of production of final goods and services using current market prices to determine the value of each unit that is produced. Real GDP is the total value of production of final goods and services using market prices from a specific base year to determine the value of each unit that is produced.

Real GDP is the sum of the market value using prices from a base year. This base year may differ from the year in which the quantities were actually produce. For example, if we use the prices in 2019 as the base year in computing 2020 GDP in the example above, then the GDP would only be $300,000. I see these computations below:

$30,000 (2019 price) * 10 (2020 production quantity) = $300,000 (2020 GDP)

By holding price is constant, using prices from a single base year such as 2019, we are able to make meaningful comparisons across the years. Economists say that such analyses use constant dollars. In this case, the constant dollars are based on prices from 2019, therefore we are using constant 2019 dollars. We can compute real GDP growth by comparing the real GDP across two years. For example, let's say in 2020 that the production of vehicles increases to 12 vehicles. Using 2019 prices, we can compute the real GDP in 2020:

$30,000 (2019 price) * 12 (2020 production quantity) = $360,000 (2020 GDP)

No, we can compute the growth in real GDP:

$$\frac{\$360,000 - \$300,000}{\$300,000}$$

$$\frac{\$60,000}{\$300,000}$$

$$0.2$$

Therefore, real GDP has grown by 20%. In reality, most countries do not experience this kind of insane gross. Real GDP growth in the United States average is 3.2% per year.

The GDP Deflator is 100 times the ratio of nominal GDP to real GDP in the same year. It is a measure of how prices of goods and services produced in a country have risen since the base year.

Let's use our Ford motor vehicles as an example. If we are looking to compute the GDP deflator for 2020 with a production of 12 vehicles and 2019 base year prices. Let's say the price of vehicles in 2019 was $30,000 per vehicle, and the price of vehicles in 2020 was $40,000 per vehicle. Ford motor vehicles produced 10 vehicles in 2019 and 12 vehicles in 2020. Knowing this information, the nominal GDP of 2020 is $480,000 (12*$40,000) and the real GDP of 2020 is $360,000.

$$GDPDeflatorof2020 = \frac{NominalGDPof2020}{RealGDPof2020} * 100$$
$$GDPDeflatorof2020 = \frac{480,000}{360,000} * 100$$
$$GDPDeflatorof2020 = 1.33 * 100$$
$$GDPDeflatorof2020 = 133$$

Therefore, this can be interpreted as "Prices have risen 33% since 2019." The GDP deflator is an excellent method to measure price movements.

Another method to measure price movements is the consumer price index (CPI). The consumer price index is 100 times the ratio of the cost of buying a basket of consumer goods using target your prices divided by the cost of buying the same basket of consumer goods using base year prices. The formal equation is below:

CPI (target year) = (cost of buying a particular basket of consumer goods using target year prices)/(cost of buying particular basket of consumer goods using base year prices) * 100

The GDP deflator and the CPI are almost the same. The only difference is that the GDP deflator studies the basket of goods that is produced domestically. This is the GDP basket. The CPI studies a particular basket of consumer goods. This is a consumer basket. There are three key differences between the GDP basket and the consumer basket:

1. The GDP basket includes things that households don't always purchase, like coal fired power plants and subway stations. Although consumers use the services that are provided by governments, no consumer purchases a subway directly. Therefore, they appear in the G key basket but not in the consumer basket.

2. The consumer basket includes things that households purchase but are not counted in the GDP. For example, the GDP counts only domestic production, so it does not count imports, such as the foreign value added in a laptop manufactured abroad.

3. Even if a product isn't included in the GDP and the consumer basket, it is likely to have a different weight in each basket. For example, housing related expenditures are included in both baskets, but housing has a larger role in the consumer basket. Housing represents more than 40% of the consumer basket, but only 20% of the GDP basket. This is because housing represents a huge chunk of household expenditures but in the grand scheme of the nation, only represents 20% of expenditures.

The rate of increase in prices is the inflation rate. It is calculated as a year over year percentage increase in a price index. The price index is either the GDP deflator or the CPI.

3.2 Aggregate Incomes

Inequality Around the World

Income varies across countries. In order to understand why it does vary, we need to understand some macroeconomic metrics to evaluate these differences. In the recent chapter, we learned about the national income accounting identity. We can compute the national income by evaluating production, expenditure or income. All three of which give us the same metric, and that is the gross domestic product, the GDP. The GDP per capita is the GDP divided by the total population. It's also referred to as income per capita. The formula to compute this is as follows:

GDP per capita = GDP/total population

For example, in 2018, the United States had a nominal GDP of $20.54 trillion. The population was 327 million people. Therefore, the nominal GDP per capita was approximately $62,813. In comparison to Mexico, which had a GDP per capita of $9,640 in the same year, after converting pesos to US dollars. We can compare this GDP per capita to that of Germany's, which was $47,603 in 2018. We can already see the income inequality between the three countries. However, we need to learn about purchasing power parity to evaluate this inequality with more depth. The **purchasing power parity** constructs the cost of a representative basket of commodities in each country and uses these relative costs for comparing income across countries. For example, the cost of a phone call in Mexico is different from the cost of a phone call in the United States. The purchasing power parity takes into account the differences in these prices. The purchasing power parity provides a better way to convert GDP in domestic currencies into common units. For example, a change in the US Mexico exchange rate would consequently lead to a different Mexico GDP calculation. For example, the exchange rate between the peso in the dollar was 19.54 in January 2019. If we had instead used the exchange rate in January 2017, which was 18.93 pesos per dollar, then the average GDP (GDP per capita) in Mexico would have been $9,984 instead of $9,640. However, this fluctuation has little to do with the prices that households pay for goods and services in Mexico or the United States. This is another reason why we study the purchasing power parity (PPP) concept. Once we have constructed the purchasing power parity, we can then measure a country's GDP in PPP adjusted US dollars. For example, this representative basket costs $1 in the United States and 9.23 pesos in Mexico in 2019. On this basis, the PPP conversion factor between US dollars and pesos is $1 for 9.23 pesos. Using this procedure, the GDP per capita in Mexico and PPP can be compared by multiplying the GDP per capita in Mexico in pesos by the peso dollar PPP conversion factor that we just derived:

Mexico GDP per capita in PPP $ = Mexico GDP per capita in pesos * ($/ peso PPP)

Mexico GDP per capita in PPP $ = 189,010 * ($/ peso PPP)

Mexico GDP per capita in PPP $ = 20,478

There is a significant difference between exchange rate based measures and PPP based measures of GDP per capita. For example, the difference between the US economy and poorer economies is generally smaller when we use PPP-based measures. This pattern reflects the lower cost of living in countries with lower GDP per capita. In other words, the exchange rate based

comparisons of GDP ignore the fact that many commodities are cheaper in poorer countries.

GDP per Worker

The GDP per worker is defined as the GDP divided by the number of people in employment. This is different from the GDP per capita. The GDP per capita takes into account the entire population of the country. GDP per worker only takes into account the population of employed people. Therefore, the GDP per worker is typically higher than the GDP per capita. This is because the GDP per capita typically has a larger denominator. For example, in Mexico, the PPP-adjusted GDP per capita in 2018 given 2011 constant dollars was $18,133. The PPP-adjusted GDP per worker for Mexico in 2018, again using 2011 constant dollars, was $40,453. The primary reason to explain the difference in GDP per capita or GDP per worker between countries is the varying productivity of each country. Productivity refers to the value of goods and services that a worker generates for each hour of work. For example, some countries are more productive in producing goods and services compared to others. To understand why some countries are more productive than others, we need to study the factors that make labor more productive in some countries compared to others.

Although the GDP per capita is an important metric to understand, it does not tell us everything about a country's well-being. For example, it does not give us a good idea of the wealth distribution. In the United States, in 2015, the top 1% of the wealthiest people owned 22% of all of the income in the United States. In Mexico, in the same year, the wealthiest 1% own 13.6% of all the income in Mexico. This kind of wealth distribution offers rich information, that is not available from just looking at the GDP per capita. GDP also leaves out information such as pollution, quality of healthcare, and public safety.

One thing that the real GDP per capita DOES indicate is the number of people living in extreme poverty in that country. The world bank has established that absolute poverty is defined as living on less than $1.08 per day. It is a measure of absolute poverty used by economists and other social scientists to compare the extent of poverty across countries. We can see this relationship in the graph below. The real GDP per capita is on the X axis and a percentage of that countries population living in absolute poverty is on the Y axis. Just from a visual analysis, we can see that countries with a higher GDP per capita have a lower percentage of the population living in absolute poverty. This graph shows a strong association, indicating that GDP per capita gives us a fairly good idea of which countries have populations suffering from extreme poverty. In this graph, we are using a proportional scale, which stretches the X axis so that a 10% change in GDP per capita represents the same absolute distance on the horizontal scale, regardless of whether we are starting from a lower level of $500 or a higher level of $800.

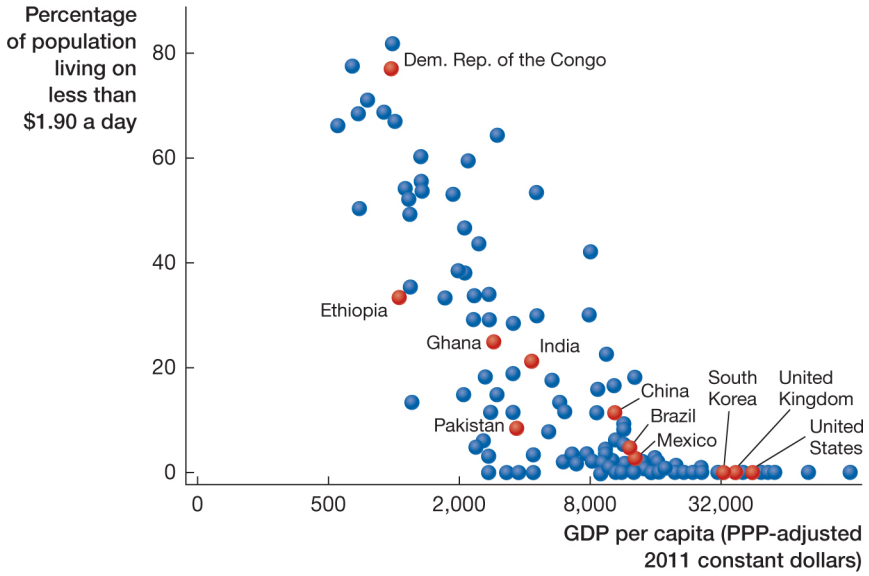

Figure 3.2: The relationship between poverty and GDP per capita in 2014 using PPP adjusted 2011 constant dollars.

Productivity and the Aggregate Production Function

There are three primary reasons to explain the different productivity levels across countries:

1. **Human Capital**

   Human capital is each person's stock of skills to produce output or economic value. For example, a worker with a degree in computer science will be much more productive in computer programming compared to a worker with a degree in environmental science. The person with the computer science degree is more productive in writing code compared to the person with the environmental science degree.

2. **Physical Capital**

   Physical capital is the stock of all machines and buildings used for production. For example, in agriculture, agricultural machinery contributes to aggregate production. This includes the equipment used for transporting inputs and outputs, the buildings used to store output, and other things. Physical capital stock of an economy is the value of equipment, structures, and other nonlabor inputs used in production. Workers will be more productive when the economy has a bigger physical capital stock, enabling each worker to work with more equipment.

3. **Technology**

   Technology refers to a set of devices and practices that determine how efficiently and economy uses its labor and capital. An economy with better technology uses its labor and capital more efficiently and therefore achieves higher productivity. An economy can achieve better technology by either using superior knowledge in production or organizing production more efficiently.

The Aggregate Production Function

The aggregate production function describes the relationship between the aggregate GDP of a nation and its factors of production. The factors of production are the inputs that we used in the production process to produce goods or services. To understand a nation's output, we will look at a production function that describes how the factors of production are combined to produce GDP. We are not focusing on specific commodities here, we are focusing on the overall production of all goods and services. This is the aggregate production function.

The factors of production include labor, physical capital in land and technology.

1. **Labor**

   Labor is the most important factor of production. A nation can increase output by employing more workers. For example, more workers can be deployed for tilling the soil and harvesting corn, which would therefore increase the production of total corn. Although increasing the labor supply is important, not all workers are the same. Some workers have greater human capital, and therefore greater productivity, than others. Therefore, looking at the total number of workers in an economy can be misleading. We need to know the total efficiency units of labor. The total efficiency units of labor includes the product of the total number of workers in the economy and the average human capital of workers. For example, a computer science undergraduate student can produce twice the amount of Python scripts compared to a high school student. The formal definition of the total efficiency units of labor is below, Where L is the total number of workers in the economy and lowercase H is the average efficiency or human capital of workers. Acquiring more skills through formal schooling is one way for a worker to increase their productivity.

   $$H = L * h$$

2. **Physical Capital and Land**

   The second most important factor of production is physical capital. This includes machinery, equipment and structures. When economy has more physical capital, or a greater physical capital stock, its workers can work with better equipment and structures and the economy will produce more GDP. Land is also a factor of production. The value of land and natural resources can be included in physical capital stock.

3. **Technology**

   Technology determines how efficiently the economy uses its inputs, (labor, capital and land.) In the aggregate production function, technology summarizes the relationship between the factors of production and GDP. Better technology means that the economy can generate more output from the same set of inputs, and thus increases its productivity for a given number of efficiency units of labor and capital.

We can put all of these factor inputs together to form the aggregate production function. The aggregate production function is as follows:

$$Y = A * F(K,H)$$

Where:

1. Y stands for GDP
2. K is the physical capital stock of the nation
3. H the efficiency units of labor that the economy uses in production
4. The function F signifies that there is a relationship between physical capital, labor, and GDP. In particular, GDP is generated through a combination of physical capital and the efficiency units of labor.
5. A is an index of technology. As a increases, the economy produces more GDP with the same level of physical capital stock in total efficiency units of labor.

According to the aggregate production function, a change in either labor or physical capital can increase GDP. If labor is held constant, in other words if the labor force does not increase or decrease across years, but the physical capital stock increases, then we can produce more GDP. If physical capital is held constant, but labor increases, we can produce more GDP. However, technology is the most important factor of production in this function.

It is important to note that although the aggregate production function is a wonderful metric, it is subject to the law of diminishing marginal product. The law of Diminishing marginal product states that the marginal contribution of a factor of production to GDP diminishes when we increase the quantity used of that factor of production. This means that although each extra worker increases the overall GDP, they do so at a decreasing rate. For example, the first worker may produce $10 in GDP, the second worker will produce $8, and the third worker will produce $6. Although each worker is still contributing to an increase in GDP, they are doing so at a decreasing rate.

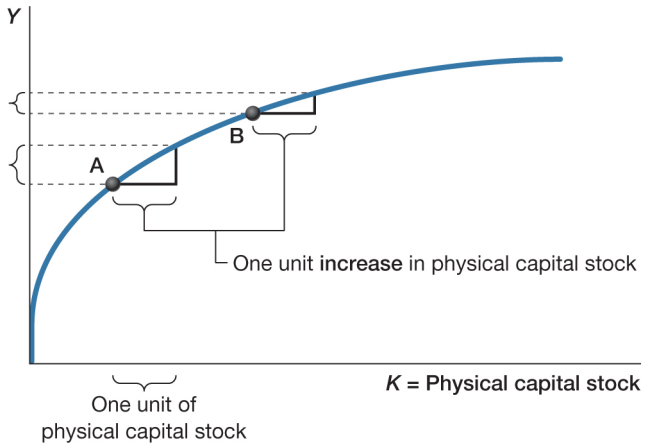

Figure 3.3: the aggregate production function with physical capital stock on the X axis

The graph above shows the positive relationship between physical capital and output, as well as the law of diminishing marginal returns. This is demonstrated by the stagnating curve. The first few units of physical capital is significantly increase overall output, and then the curve flattens out.

The Role and Determinants of Technology

Technology determines how efficiently and economies inputs are put to use. When technology improves, the relationship between GDP and the physical capital stock shifts up. We can see this in the graph below. Therefore, for every level of the efficiency units of labor, a better technology implies the economy will then produce more GDP.

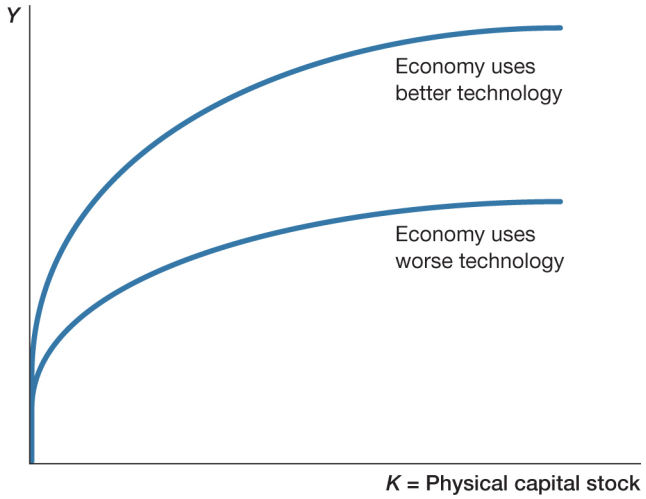

Figure 3.4: The shift in the production function resulting from more advanced technology

There are two important dimensions of technology to understand: knowledge and efficiency of production. Knowledge encompasses the idea that we know how to produce many new goods which we did not know how to previously produce. In addition, knowledge enables us to perform certain tasks more efficiently. Part of this knowledge is the human capital of workers: workers today can perform a wide range of tasks more productively compared to our grandparents. Technology can include certain types of physical capital. For example, computers are a physical capital that also represent a new form of technology that helps workers to do their tasks more efficiently. Advances in technology typically come from research and development initiatives. This includes a wide range of activities such as research at universities in private labs, or research in supply chain management production to improve the efficiency of a factory. Research and development is a major activity in the United States economy. Around one. An estimated 2.5 million people worked as researchers in 2012 and $457 billion we spent on research and development. Almost half of this amount was spent by businesses, and they remaining portion was spent by public institutions and government bureaucracies.

The second dimension of technology is the efficiency of production. The efficiency of production refers to the ability of an economy to produce the maximum amount of output from a given amount of factors of production and knowledge. For example, let's consider two economies. In

one economy, the college graduates with an economics degree go on to teach economics at universities. The college graduates with an environmental science degree go on to perform environmental research. Now let's consider economy #2. In economy #2, the economic graduates go on to do environment research and the environmental graduates go on to teach economics. Although both economies have the same amount of human capital, they are allocating their resources differently. The first economy will be much more successful in both teaching economics and doing environmental research. The difference Between the two economies lies in their efficiency of production. Economy #2 is inefficient in their resource allocations compared to economy #1. When the economy is able to increase the efficiency of production, there will be a shift in the aggregate production function. Therefore, the efficiency of production is included in the definition of the aggregate production function. The importance of technology for GDP is the reason why we include A in the aggregate production function.

$$Y = A * F(K,H)$$

4 Week 4

4.1 Economic Growth

The United States real Gross Domestic Product (GDP) per capita today is about 25x what it was in 1820. In the 1800s, only a small fraction of the population lived in urbanized areas, most individuals worked in agriculture. The United States has experienced significant economic growth over the past 200 years. This has contributed to the growth in real GDP per capita. In this unit, we'll address how and why we have achieved this kind of significant economic growth.

Quick Review from Previous Topics: Real vs. Nominal GDP

Real GDP is the total value of production (final goods and services), using market prices from a specific base year to determine the value of each unit that is produced.

Nominal GDP is the total value of production (final goods and services), using current market prices to determine the value of each unit that is produced.

Real GDP is the GDP of a country after being adjusted for inflation. For example, if a country's GDP is $100 in 2001 and $500 in 2021, this could be due to inflation. This increase in GDP may not accurately reflect economic growth. Therefore, we adjust for inflation to get a more "true" estimate of the GDP. After adjusting for inflation, the real GDP is typically lower than the nominal GDP. The inflated GDP value is also referred to as "nominal GDP." To compute this year's real GDP, we multiply this year's production by the base year's prices. The base year is typically a year way back in the past, as designated by the Bureau of Economic Analysis. For example, the currently designated base year is 2012.

The GDP Deflator is 100 times the ratio of nominal GDP to real GDP in the same year. It is a measure of how prices of goods and services produced in a country have risen since the base year. For example, computing the GDP deflator of 2020, assuming a base year of 2009, looks like this:

$$GDPDeflator of 2020 = \frac{Nominal GDP of 2020}{Real GDP of 2020} * 100$$

$$GDPDeflator of 2020 = \frac{900,000}{700,000} * 100$$

$$GDPDeflator of 2020 = 128.6$$

Therefore, this can be interpreted as "Prices have increased by 28.6% since 2009." The "base year" used in adjusting for inflation is commonly referred to as "constant dollars."

The Power of Economic Growth

The U.S. economy has significantly grown from 1820 to 2016. Economists like to measure this in terms of real GDP per capita. The "per capita" metric of something is measured as the total amount (of whatever we are measuring) divided by the total population. For example, the GDP per capita in the United States is computed as:

$$Real GDP per Capita = \frac{Total Real GDP}{Population}$$

Economic growth is the increase in real GDP per capita of an economy. As we can tell from the graph below, the growth in real GDP has not been constant. We can see that there has been an increase in GDP overall, but there have been several downward trends, such as the Great Depression in the 1930s. These downward trends are referred to as "contractions." However, these contractions are typically temporary, as was the Great Depression.

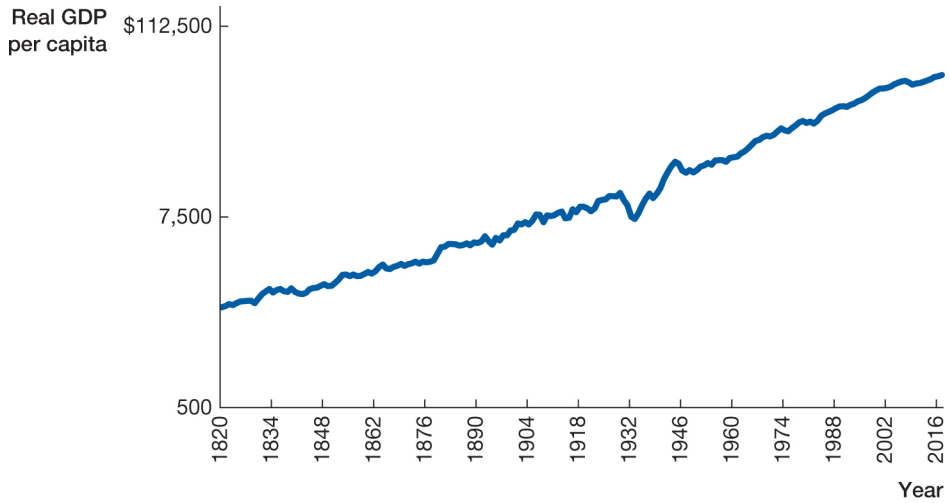

Figure 4.1: Real GDP per Capita in the United States (2011 Constant Dollars)

Real GDP per capita has increased significantly. For example, it was $2,806 in 1820 compared to $14,655 in 1950 and $50,752 in 2014. These values are all in 2011 constant dollars.

A growth rate is the change in quantity between two dates, relative to the baseline quantity (quantity at the initial date). This is the formula to compute a growth rate (in percent):

$$Growth_{t,t+1} = \frac{y_{t+1} - y_t}{y_t} * 100$$

Where:

- y_{t+1} = GDP in next period
- y_t = GDP in previous period

Let's do an example and compute the growth rate of the real GDP per capita from 2005 to 2006.

$$Growth_{2005,2006} = \frac{\$51,374 - \$50,512}{\$50,512} * 100$$
$$Growth_{2005,2006} = 0.017 * 100$$
$$Growth_{2005,2006} = 1.7\%$$

Therefore, the real GDP per capita grew by 1.7% from 2005 to 2006. It is important to note that the annual growth rate of real GDP per capita can be positive over a period, even if there are some economic contractions in that period. For example, we can see this in the graph below, the annual growth rate of real GDP per capita from 1950 to 2014 was 2.03%. However, there were several economic fluctuations in that time period. This 2.03% represents the average annual growth rate in those 64 years.

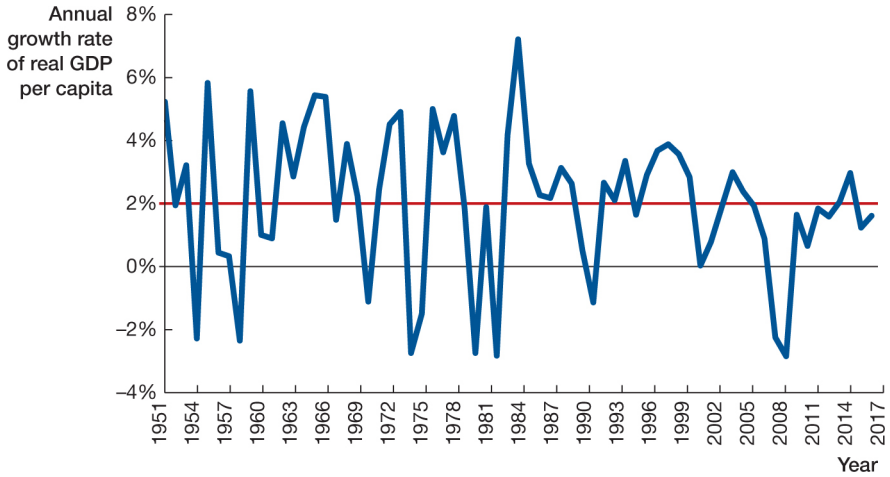

Figure 4.2: The Annual Growth Rate of Real GDP per Capita in the United States Between 1950 and 2014 (2011 Constant Dollars)

Now that we understand the concept of a growth rate, we can understand the idea of exponential growth. **Exponential growth** is where the growth process can be described by a constant growth rate. More importantly, exponential growth is the idea that new growth builds on past growth, with compounding effects. This is one reason why we see such a vast difference in GDP per capita values across countries. Some countries experience exponential growth but others do not. Exponential growth is computed differently from just a straightforward growth rate. For example, the formula that we use to compute the new GDP given an exponential growth rate is the following:

$$NewGDP = (initialGDP) * (1 + g)^n$$

Where:

- g = constant growth rate
- n = number of years

For example, let's compute the GDP after 15 years, assuming a constant growth rate of 5%, and an initial starting GDP of $2.

$$NewGDP = (2)(1 + 0.05)^{15}$$
$$NewGDP = (2)(1.05)^{15}$$
$$NewGDP = 2 * 2.0789$$
$$NewGDP = 4.1578$$

Therefore, the GDP after 15 years, with a constant growth rate of 5%, will have increased by 2.0789. Then we multiply this value and the initial starting GDP of 2.0, to get the new GDP after 15 years of 4.1578.

I want to show you why the exponential growth rate is so important to understand, and why countries benefit so much from this idea of exponential growth. Let's do an example, using the same formula, and compute the exponential growth after two individual years (from year 1 to year 2, and from year 2 to year 3), given a constant growth rate of 5%, and a initial starting GDP in 2.0.

$$NewGDP = (2)(1 + 0.05)^1$$
$$NewGDP = (2)(1.05)^1$$
$$NewGDP = 2 * 1.05$$
$$NewGDP = 2.10$$

Year 2 GDP is 2.10. Now we'll compute the new GDP after the second year. This time, our "initial GDP" is 2.10. This is because we are now measuring the growth from year 2 to year 3. Year 2 is now the initial year.

$$Year3GDP = (2.10)(1 + 0.05)^1$$

$$Year3GDP = (2.10)(1.05)^1$$

$$Year3GDP = 2.10 * 1.05$$

$$Year3GDP = 2.205$$

We can see that from Year 2 to Year 3, GDP increased by 2.205 - 2.10 = 0.105. From Year 1 to Year 2, GDP only increased by 2.10 - 2.0 = 0.100. Therefore, even though both increases correspond to a 5% growth rate, there was a compounding effect resulting in a higher increase from Year 2 to Year 3 compared to the increase from Year 1 to Year 2.

When there is exponential growth, it is much better to represent this on a graph using a proportional scale. The figure on the top is using a proportional scale whereas the figure on the bottom is using a nonproportional scale.

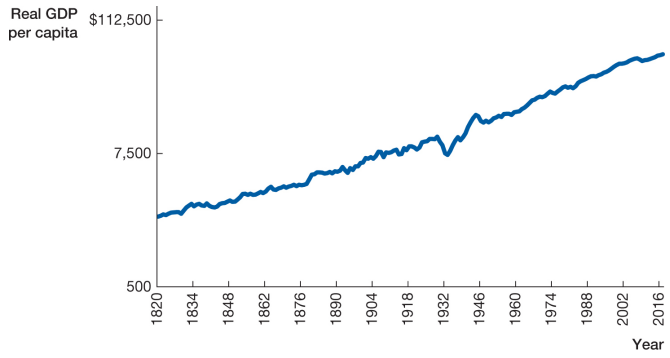

Figure 4.3: Real GDP per Capita in the United States (2011 Constant Dollars) (Proportional Scale)

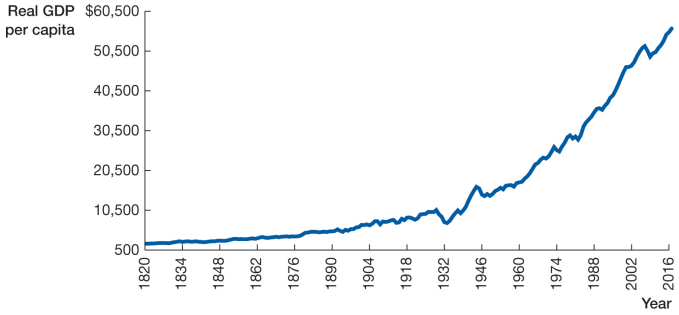

Figure 4.4: The Annual Growth Rate of Real GDP per Capita in the United States Between 1950 and 2014 (2011 Constant Dollars) (Nonproportional Scale)

A constant growth rate of 10% with an initial value of 1000 would yield a final value of 1100. However, a constant growth rate of 10% with an initial value of 100,000 would you hold a final value of 110,000. Those two increments are vastly different, one is 100, the other is 10,000, however they are the same with regards to the proportion of the base value, which is 10%. Therefore, the proportional scale on a graph shows this change where the constant growth rate of 10% corresponds to the same distance on the Y axis regardless of whether we start at an initial value of 1000 or an initial value of 100,000. Therefore, using a non-proportional scale could be misleading. You can see in the non-proportional scale figure that it seems as though the GDP is accelerating, even though the growth rate is constant.

It is important to understand the general patterns of GDP grows across several countries. From looking at the table below, we can compare GDP growth for various countries. This year should be "2014" instead of "2017." The implied annual growth rate shows how much on average each country needed to grow each year in order to reach the 2014 level given the starting 1960 level.

Country	Real GDP per Capita		Implied (Average) Annual Growth
	1960	2017	
United States	$17,600.11	$56,153.42	2.07%
United Kingdom	$11,959.49	$42,137.82	2.25%
France	$10,465.52	$40,975.05	2.43%
Mexico	$5,741.75	$18,360.42	2.01%
Spain	$5,741.40	$37,232.80	3.35%
Nicaragua	$4,476.47	$5,360.22	0.32%
Ghana	$2,816.50	$5,153.55	0.37%
Singapore	$2,663.43	$79,842.57	6.16%
Brazil	$2,463.11	$14,108.92	3.29%
Democratic Republic of the Congo	$2,422.75	$798.68	−2.03%
Guatemala	$2,418.48	$7,473.34	2.06%
Kenya	$1,749.13	$2,987.50	0.94%
South Korea	$1,175.10	$37,725.07	6.37%
China	$1,154.19	$13,051.32	4.56%
India	$1,033.67	$6,281.54	3.23%
Rwanda	$962.58	$1,948.49	1.24%
Botswana	$427.35	$16,235.75	6.60%

Figure 4.5: GDP per capita and growth in selected countries parentheses PPP adjusted 2011 constant dollars

In this case PPP stands for "purchasing power parity." Purchasing power parity is the application of evaluating GDP in each country by looking at a specific basket of goods to get a fair comparison of the cost of living standards across countries. The implied annual growth rate in the third column shows how much on average each country needed to grow each year in order to reach the 2014 level given the starting 1960 number. From these comparisons, we can see how countries stack up against each other. For example, both the United States and the United Kingdom show an average annual growth rate of about 2% in this time period. There's also been an even greater increase in GDP per capita for Singapore, Spain, South Korea, Botswana, and China. All five of

these countries were significantly poor in the United States in 1960, but they closed the gap by 2014.

We can also see that other nations who did not fare well in closing this wealth gap, such as Mexico, Brazil and India. Other countries, such as Guatemala, Kenya, Ghana, Rwanda, and Nicaragua have become relatively poor over the same time.

We can see that Nicaragua's GDP actually declined at a rate of 0.32% per year, while the democratic republic of Congo has seen a decline of 2.03% per year. As a result, both countries are worse off in 2014 compared to 1960, and this is mostly due to the decades of Civil War and political turmoil. We can see that Nicaragua's GDP is increasing at a decreasing rate, whereas the Democratic republic of Congo's GDP decreased overall in that same time.

Catch-up growth refers to the idea where a relatively poor nation increases its income by taking advantage of knowledge and technologies already invented by other countries, that are more advanced. These countries are trying to "catch up" to other countries, by benefiting from their technological advances. Japan is a good example of "catch up" growth because Japan's initial growth was very high. During the period between 1911 to 1940, Japan was the fastest-growing economy in the world. It invested heavily. After the Second World War, however, Japan's economy was destroyed. The country rebuilt a sustainable environment for economic growth during the 1950s and began importing machinery and technology from the United States. It clocked incredible growth rates in the period between 1960 to the early 1980s.

Even as Japan's economy powered forward, the United States' economy, which was a source for much of Japan's infrastructural and industrial underpinnings, hummed along. Then by the late 1970s, when the Japanese economy ranked among the world's top five, its growth rate had slowed down. It saw high growth in GDP as it "caught up" to other nations. However, this declined in later years once Japan "caught up" to other countries.

Sustained growth refers to the idea where a country's real GDP per capita grows at a positive and steady rate for long periods of time. The U.S. and the U.K. are good examples of sustained growth because they have maintained positive, steady rates of growth.

What fuels economic growth?

In order to understand the drivers behind economic growth, we first need to understand the aggregate production function. This function links the two factors of production, which is physical capital (K) and efficiency units of labor (H). It also takes into account a country's technology level (A). Technology captures the level of productivity that comes from innovations as well as the efficiency of production. When technology improves, the aggregate production function increases. This is because, if we improve our technology, but are using the same levels of physical capital and labor, we'll be able to produce more overall, just because the technology has improved.

$$Y = A * F(K, H)$$

The stock of physical capital in a country includes the value of all equipment, such as machines, cars, planes, and computers. It also includes the value of structures, like buildings. This physical capital stock can be increased by investment, known as physical capital accumulation. This is the basic idea that after building one structure, you can build the next. You don't have to start "from scratch." Therefore, you can accumulate capital, and your overall value of capital stock will build overtime.

The national income accounting identity is computed as the following:

$$Y = C + I + G + X - M$$

Where:

- Y = GDP
- C = consumption (household expenditures on consumption of goods and services)
- I = investment (expenditures on investment goods by private agents)
- G = government purchases of goods and services
- X = exports
- M = imports

In a closed economy, there are no exports or imports, therefore X = M = 0. Another way to interpret investment in a closed economy is to understand that investment comes directly from savings. In a closed economy, all income is either consumed or saved. Therefore, any investments would come from savings. We can further interpret this as, household savings will be then allocated to firms that will use them for investments. For example, banks take money deposit by households and lend this out to firms for investments. Consequently, nations with high overall savings rates can accumulate and invest in physical capital rapidly, and therefore increase in the aggregate production function, therefore increasing GDP.

The Importance of Savings

The national savings rate plays a huge role in economic growth. Although the government can invest its tax revenues into public infrastructures, the resources for these investments come from household savings. To understand the national savings rate, we need to understand the individual decisions that each household makes in choosing between consumption and savings. This is an important optimization decision to understand. Each household faces different priorities that influences decisions to consume income today versus save it for tomorrow. As with all optimization problems, these choices are affected by prices. In this case, the relevant price of savings is the interest rate. The interest rate determines the rate of return that a household expects on their savings.

Higher interest rates typically encourage higher savings. Also, expectations of future income growth will have an impact on the saving decision. For example, households that expect a significant increase in income in the future may have less reason to save money today. This is because, if they have optimistic expectations, they may be less likely to save for a "rainy day" today. These kind of trades determine the saving rate of an economy. The **saving rate** designates the fraction of income that is put towards savings.

We can compute the saving rate of an economy as the quotient of total saving and GDP. For example, the total savings in the economy in 2013 was 2.18 trillion, well the GDP was 16.8 trillion. Therefore, the saving rate was 12.98%.

$$SavingRate = \frac{Totalsaving}{GDP}$$

$$SavingRate = \frac{2.18trillion}{16.8trillion}$$

$$SavingRate = 0.1298$$

What contributes to sustained growth?

Technology is the primary driver of sustained growth. There is one reason why physical capital and labor cannot contribute to sustained growth, and that is the idea of diminishing marginal returns. Diminishing marginal returns is the idea that with every incremental increase in physical capital, comes a slightly smaller increase in overall GDP. Therefore, although more physical capital results in a higher GDP, it comes at a slower rate. The same goes for labor, although each new worker results in a higher output, it comes at a slower rate. This is very similar to the idea of "too many cooks in the kitchen." Those capital and labor exhibits diminishing marginal returns.

Technological change is the process of new technologies, goods, and services being invented, and using the economy. This enables the economy to achieve a higher level of real GDP given the physical capital and labor. Some examples of this might be the improvement in the computational power of computers, or the advent of the transistor, or the improvement in lighting technologies. Technological change can contribute to sustained growth because it occurs at a constant rate, rather than constant increments. New technology grows at an exponential rate, because it builds on previous technology. Therefore, technology does not suffer from diminishing marginal returns like physical capital and labor do.

The Malthusian Model

The Malthusian model is a dismal one, but it is one type of population growth model. The theory rests on the idea that food is necessary for subsistence, and that the world cannot simply provide enough food for its growing population. Thomas Malthus was a British philosopher, who

was active in producing work in the late 1700s. Malthus argued that when the standards of living improved, couples within produce more children, increasing the fertility rate of a nation, but because real GDP cannot grow as fast as a population, this will then put downward pressures on the real GDP per capita. This will then contribute to a fall in real GDP per capita, which would trigger famines, wars, and overall higher mortality. He referred to this process as the **Malthusian cycle**. Although this seems ludicrous to an individual in the current century, this was actually a good representation of the world before 1800.

Around this time, fertility began to decline. We saw this primarily as countries transitioned from an agriculture based to an industry based economy. Families had fewer children, and societies underwent a process known as demographic transition. **Demographic transition** refers to the overall decline in fertility and number of children per family. Until the demographic transition in the nineteenth century, there were recurring Malthusian cycles. Many historians view this transition as a key ingredient in modern growth because it enables economies that experienced reduced fertility to break out of this Malthusian cycle and instead experience economic growth.

The Industrial Revolution

The **Industrial Revolution** was a gradual process. It was a series of innovations implemented in the production process that begin at the end of the 18th century in Britain. It is the start of the arrival of new machines and methods of production, starting in textile manufacturing and spreading to other sectors. The industrial revolution is important because it is a starting point of industrialization at that spread to many other countries. Countries that are rich today are those that have managed to achieve steady growth rate of the past 200 years. They are also the ones that benefited from the technologies brought about by this industrial revolution.

Many of the technologies that we take for granted today have been available to us over the past 250 years, including telecommunication, computers, electricity, etc. These advances are the result of exponential growth and technology since the Industrial Revolution. An important foundation that fuels this technological growth is investment in research and development (RD). The United States invests 2.79% of its real GDP in RD every year. Other countries invest even more, such as Israel at 4.66%, Switzerland at 3.0% and Sweden at 3.70%.

The Solow Growth Model

The Solow growth model is the primary model that economists use to formally study how GDP is determined. It is named after the economist, Robert Solow. The Solow model consists of three building blocks.

1. The aggregate production function. This function essentially links the GDP to three factors: physical capital, efficiency units of labor, and technology. Efficiency units of labor include the knowledge available to the economy and the efficiency of production, it has the potential to shift the entire aggregate production function.

$$Y = A * F(K, H)$$

2. The physical capital accumulation function. This function takes into account the value of physical capital in an economy, but also factors in the cost of depreciation, and also factors in the effect of investments.

$$K_{now} = K_{lastyear} - K_{depreciated} + I$$

$$K_{now} = K_{lastyear} - (Depreciationrate * K_{lastyear}) + I$$

$$K_{now} = (1 - d) * K_{lastyear} + I$$

Where:

- K_{now} is the physical capital stock this year.
- $K_{lastyear}$ is the physical capital stock last year.
- $(1 - d)$ is the fraction of physical capital stock that is not depreciate between the two years.
- $d * K_{lastyear}$ is the decline in the value of physical capital due to depreciation.
- I is the value of investments.

3. The household savings function. This function determines investments based on the savings rate and the total economy's output.

$$I = s * Y$$

Where:

- I is investment in the economy
- s aggregate savings rate
- Y total GDP

We can plug the aggregate production function into the household savings function and represent this in a graph.

$$I = s * A * F(K, H)$$

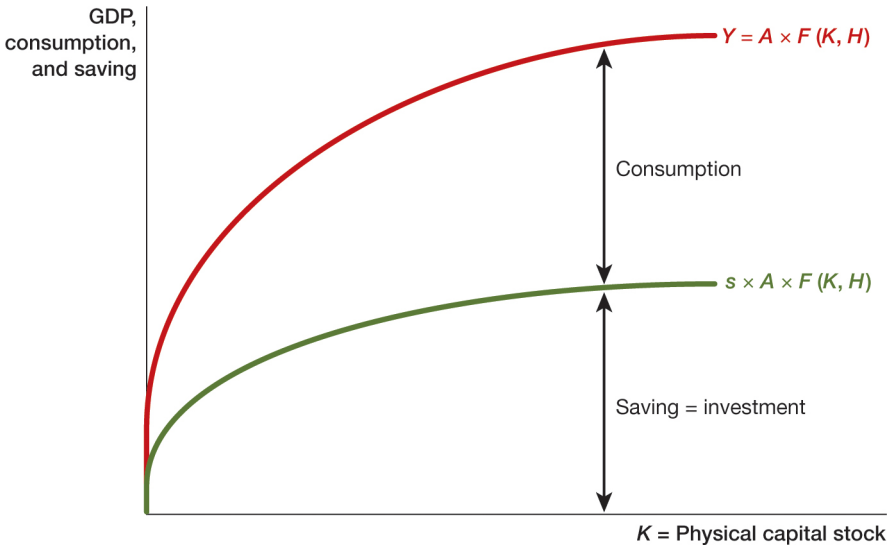

Figure 4.6: Aggregate Income and Aggregate Savings

The red curve represents the aggregate production function, which is represented by the relationship between the stock of physical capital for given levels of efficiency units of labor and technology. The green curve represents the relationship between investments and physical capital stock given the savings rate of household. It is the GDP times the savings rate. Therefore, the distance between the green curve and the X axis represents the aggregate saving or investment given a certain level of physical capital.

Steady State Equilibrium in the Solow Model

The **steady state equilibrium** is a natural situation that occurs when physical capital stock this year equals physical capital stock last year.

$$K_{now} = K_{lastyear} = K$$

We can combine this equation with the physical capital accumulation function to get the following:

$$K_{now} = (1 - d) * K_{lastyear} + I$$

$$K = (1 - d) * K + I$$
$$K = K - d * K + I$$
$$d * K = I$$

Therefore, the physical capital stock of the economy will only remain constant over time if the amount of investment (I) equals depreciated value of the physical capital stock, which is the depreciation rate of the economy(d) times the a capital stock (K).

Knowing this, we can put all of these ingredients together to determine the steady state equilibrium.

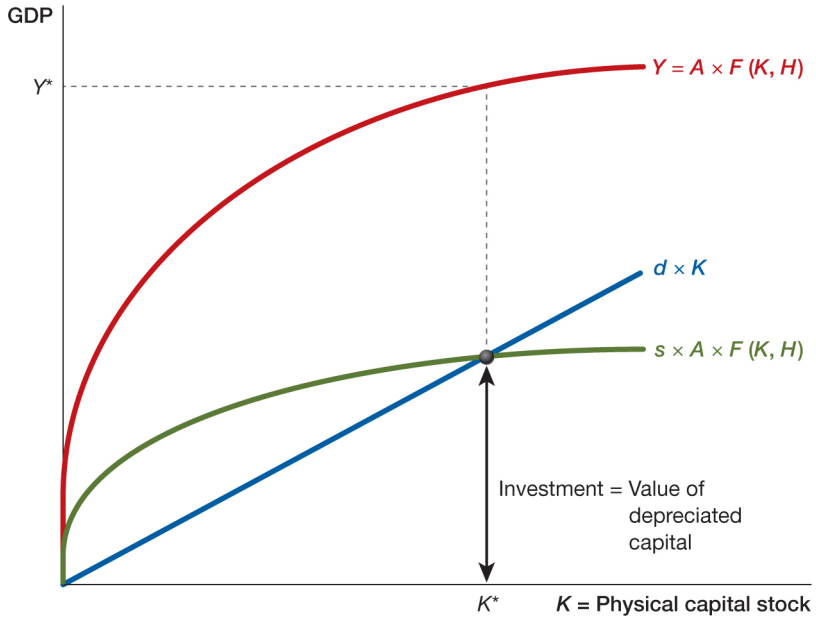

Figure 4.7: Steady State Equilibrium in the Solow Model

The steady state equilibrium is represented at the intersection of the green and blue curves. This point is the intersection of the savings curve and the investment necessary to replenish depreciated capital curve. Therefore, at the steady state equilibrium, we have a physical capital stock of K* and a GDP of Y*. Once we have the steady state equilibrium determined, we can use it to study the determinants of GDP.

Determinants of GDP

We have already discussed how the saving rate is a crucial determinant of the GDP. Let's see what happens graphically if the saving rate increases. We can compare two economies with the same aggregate production function, population and labor, but have different savings rates. In the graph below, the green line indicates a higher savings rate. The red line indicates a lower savings rate. By assumption, both economies have the same rate of depreciation. The economy with the higher savings rate will have a steady state equilibrium to the right and above the economy with

the red savings rate. An increased saving rate indicates that the physical capital stock increases to the new amount of K**, and the GDP also increases to the new amount of Y**.

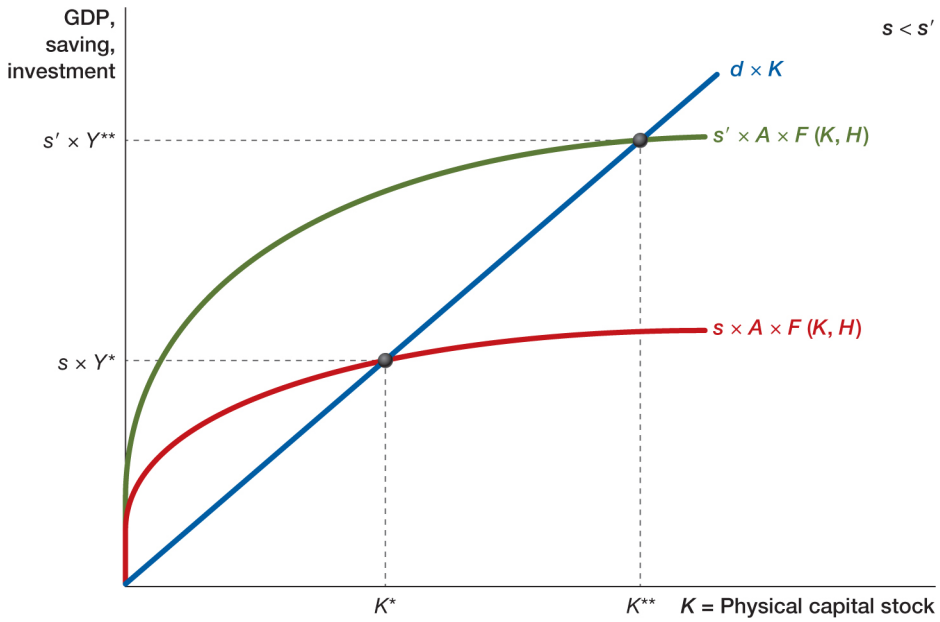

Figure 4.8: The Impact of the Saving Rate on the Steady State Equilibrium

Although a higher saving rate can increase the household saving function, household savings can also increase by an improvement in human capital or better technology, represented by A in the aggregate production function. This means that we can increase savings overall, not by actually increasing the saving rate, but by increasing our total output via improving technology or bettering human capital. If technology or human capital improves, this means that we can produce more efficiently given the same level of physical capital. Labor becomes more efficient when this happens. Therefore, the aggregate production function increases. We can see this represented by the new H_2 value in the aggregate production function in the graph below. This implies that hire human capital leads to both hire city state equilibrium physical capital stock as well as a higher real GDP for the country. Assuming there is not a change in population, then a higher real GDP translates into a higher real GDP per capita.

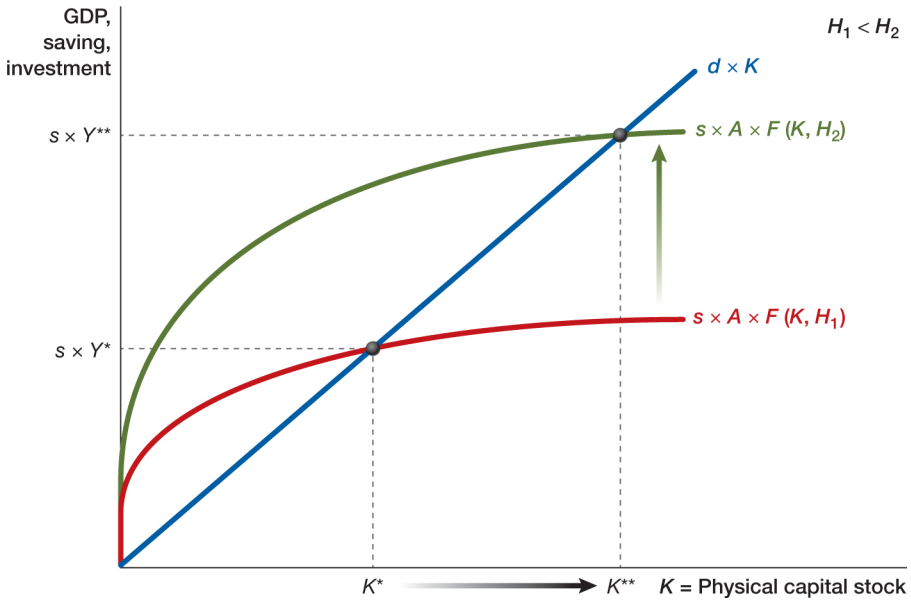

Figure 4.9: The Impact of Human Capital and Technology on the Steady State Equilibrium

It is important to remember that although savings is a crucial part of GDP, it is not responsible for sustained growth. Technology is at the root of a sustained growth. This is because there is a maximum to how much an economy can save and therefore a limit to what GDP it can achieve. For example, the maximum saving rate for an economy is 100%. However, technological improvements do not have these kind of caps. Therefore, technological improvements can provide the kind of sustained GDP growth that saving rate increases cannot. We can see this in the figures below. The top figure represents what higher saving rates can provide an economy. The bottom figure represents what technological improvements can provide an economy.

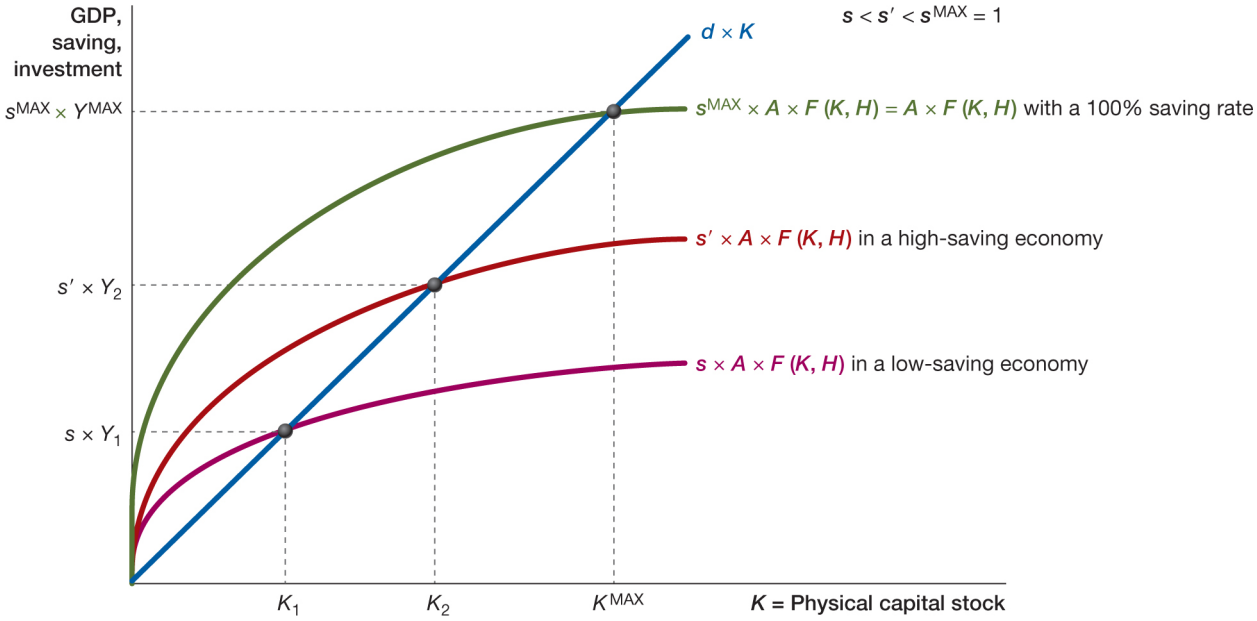

Figure 4.10: Three Economies with Different Saving Rates in the Solow Model

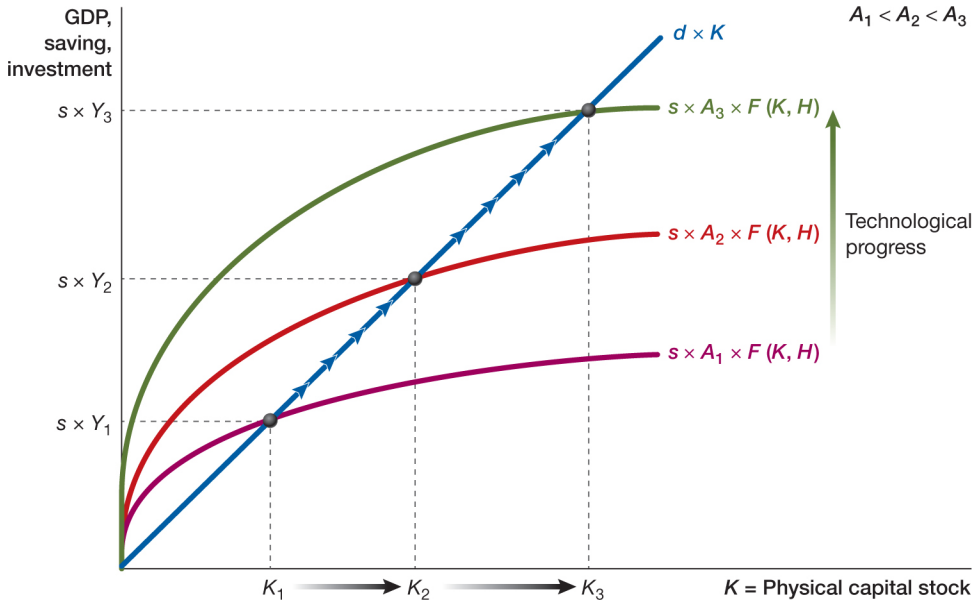

Figure 4.11: Sustained Growth Driven by Technological Change

4.2 World Development

We can see that different countries have varying PPP-adjusted GDP per capita, and in this unit, we are going to discuss why certain countries are more developed than others. One interesting thing to note is that countries around the equator tend to have reduced real PPP-adjusted GDP per capita compared to countries that are further from the equator. For example, the democratic republic of Congo, had a PPP-adjusted GDP per capita of $859. Compared to Finland, with a PPP-adjusted GDP per capita of $38,897. This has led social scientist to believe that countries around the equator are doomed to an economy of poverty.

Proximate vs. Fundamental Causes of Prosperity

The **proximate causes of prosperity** include factors such as human and physical capital, and technology. These proximate causes of prosperity would explain why certain countries have a higher real GDP per capita compared to others. However, they do not explain why these factors of production are so much higher compared to other countries. Why is human capital in Finland so much better than human capital in the Democratic Republic of Congo? Therefore, these are just the proximate causes of prosperity.

To really understand the drivers of prosperity, we need to know why certain factors of production are better in other countries compared to others. Therefore, we are looking to understand the **fundamental causes of prosperity.** These fundamental causes would explain the root cause for the disparity in the proximate causes of prosperity.

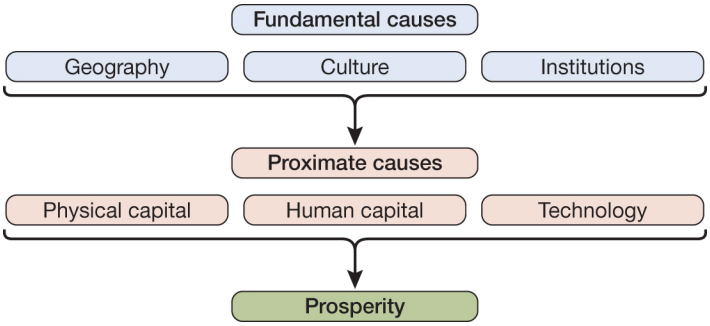

Figure 4.12: Proximate vs. Fundamental Causes of Prosperity

Societies are prosperous when they have abundant human and physical capital, however these do not explain the deep rooted factors of prosperity. Fundamental causes, such as geographic, cultural and institutional factors have an impact on prosperity.

Geography

One theory claims that geography is a fundamental cause of prosperity. The reasoning behind this is that various areas of the world experience a variety of climates due to their geography. For example, countries around the equator experience tropical weather compared to other parts of the world. According to this hypothesis, some countries have unfavorable climates that are beyond their control. For example, some countries may be in areas where the soil is inhospitable for agriculture, or lack running water to irrigate crops. If the geography of a country is a fundamental cause of prosperity, then poor nations have little reason to expect an improvement in living standards. There are also some theories that tropical zone countries suffer from the transmission of deadly diseases, such as malaria and dengue fever, which spread more easily in a tropical climate.

Culture

Another potential fundamental cause of prosperity is culture. Culture hypothesis claims that different societies respond differently to incentives due to shared experiences, religious teachings and social norms. Culture is viewed as a key determinant of the values of an economy. Therefore, it can have a significant influence on the prosperity of an economy. For example, some societies may value investment and hard work in the adoption of new technologies whereas other societies are rather suspicious of new technologies.

Institutions

Another potential fundamental cause of prosperity is institutions. The institutions hypothesis claims that the different institutions across countries contribute to how societies organize themselves and shake the incentives of its individuals and businesses. These institutions are at the root of the differences in prosperity across the world.

The best definition of institutions is from Douglass North, an economist that was awarded the Nobel prize in 1993. He defined "institutions" as the rules of the game in a society or, more for-

mally, the humanly devised constraints that shape human interaction.

By this definition, institutions are determined by the individuals of a society and can place constraints on behaviors. Institutions can shape behaviors by determining incentives. For example, institutions can place constraints on individual behaviors via formal regulations such as laws. It is illegal to sell alcohol to minors. Institutions can also influence behaviors via informal regulations, such as social norms. For example, it is socially unacceptable to attend a wedding without contributing a gift to the bride and groom.

The economy can generate a higher GDP and achieve greater prosperity if markets allocate individuals to occupations for which their productivity is the highest. Laws and regulations can encourage firms to invest in physical capital and technology. When the education system enables individuals to invest in their human capital, the productivity of an economy as a whole increases significantly. We could also see how institutions may destroy an economy, such as trade restrictions.

One interesting case study to emphasize the institutions hypothesis is the study of north and south Korea. The Korean peninsula is divided into by the 38th parallel. To the south is the Republic of Korea, also known as South Korea. South Korea has had one of the fastest growing economies in the past 60 years. To the north of the 38th parallel is the Democratic People's Republic of Korea, commonly known as North Korea. Living standards in North Korea are similar to those in sub-Saharan Africa. In 2015, the real GDP per capita in North Korea was $1,720 (in PPP-adjusted 2011 constant dollars). In the same year, the real GDP per capita of South Korea was $35,360 (in PPP-adjusted 2011 constant dollars). What explains this disparity? Geography? No, the two share similar geographies. Culture? Highly unlikely. There are very few significant differences between the cultures, certainly not before 1947 when the country was split into two. The segregation of Korea into two countries was not something that citizens willingly agreed to.

After 1947, North Korea was led by Kim Il-Sung, a leader of anti-Japanese communist partisans during World War II, and he established himself as a dictator. The country implemented a rigid form of communism, in which resources are allocated through central planning, private property is outlawed, and markets are banned. Freedoms were curtailed in the marketplace and also in every other aspect of citizens lives, except for those who remain in the elite circle around Kim Il-Sung. Today, the dictatorship continues under the rule of his grandson, Kim Jong-Un.

South Korea, on the other hand, was led by anti-communist ideals. South Korea supported market-based economics, providing incentives to businesses for investment and industrialization and investing in the education of South Koreans.

This is what we would call a "natural experiment" that allowed us to see the divergence in economic outcomes due to the differences in institutions. We can see in the graph below how economic prosperity has sharply diverged between the two countries.

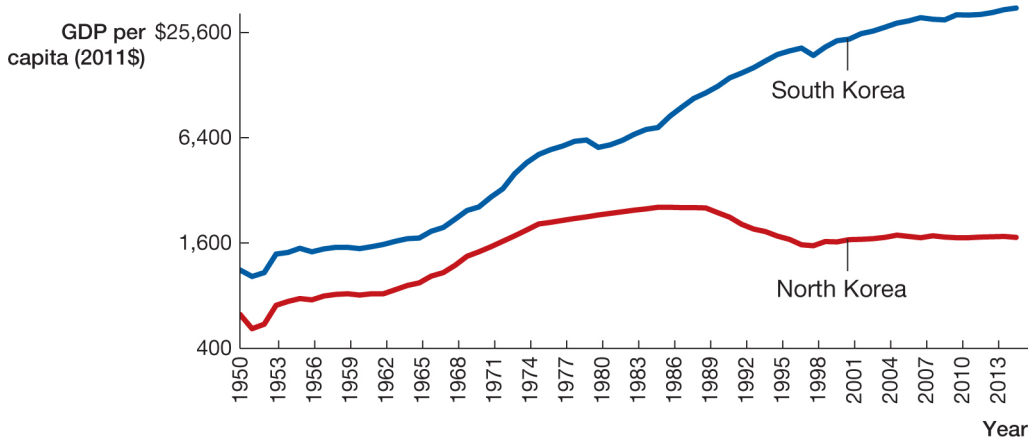

Figure 4.13: GDP per capita in North and South Korea (in PPP-adjusted 2011 constant dollars)

Institutions and Economic Development

As in the case study of north and South Korea, we can see how crucial institutions are in economic development. One reason why South Korea has such a successful economy is because of its private property rights. These **private property rights** mean that individuals can own businesses and assets and their ownership is secure. Individuals are incentivized to own property and businesses because they know the government will protect them from theft. This creates an environment that encourages development.

North and South Korea are good examples of extractive versus inclusive economic institutions. **Economic institutions** are the aspects of the societies rules that concern economic transactions. This can include the judicial system that overseas financial arrangements, how businesses can borrow money and other regulations that shape occupations and businesses.

Inclusive economic institutions protect private property, uphold law and order, allow and enforce private contracts and allow free entry into new lines of business and occupations. South Korea is a great example of an inclusive economic institution. They are inclusive because they encourage the participation of the great majority of the population in economic activities in a way that best makes use of their talents and skills.

Extractive economic institutions do not protect private property rights, do not uphold contracts and interfere with the workings of markets. They also erect significant entry barriers into businesses and occupations. The main idea of these institutions is that they work to extract resources from the rest of the society. North Korea is a great example of this. However, societies that are ruled by monarchs and dictators as well as several that hold elections for parliaments have still had extractive economic institutions. For example, some market economies with extractive economic institutions include Myanmar, Pakistan, Argentina, Guatemala, Peru, the Democratic republic of Congo, Egypt and Kenya. Even if the specific forms of these institutions differ from the extreme form of central planning communism in North Korea, they sure the fact that they failed to enforce property rights and instead privilege a few at the expense of many.

Extractive economic institutions rely on certain political institutions. Without the political elites take control of the state, North Korea would not be able to maintain its system of widespread poverty. This underscores the important role of political institutions. **Political institutions** are the aspects of the society's rules that concern the allocation of political power and the constraints on the exercise of political power. Extractive economic institutions are likely to be supported by certain types of political institutions, which concentrate political power in the hands of only a few. Inclusive economic institutions tend to coexist with different types of political institutions they distribute political power more equally in society so that no single individual can use that power for their own benefit at the expense of the rest of society.

Economic institutions can affect economic outcomes. For example, in an extractive economic institution, if a farmer who expects his output to be taken away or entirely taxed away, what incentive would he have to work? Little to none. Extractive economic institutions distort the incentives that businesses need to undertake investments and innovations.

We can see how extractive economic institutions hurt markets for entrepreneurs on the graphs below.

The return to entrepreneurship blue curve shows the number of entrepreneurs with at least the return indicated on the Y axis. It is obtained by ranking potential entrepreneurs from higher to lower return to entrepreneurship. The opportunity cost curve in red indicates the value to a potential entrepreneur of her best alternative activity. For example, if an entrepreneur chose another occupation. The intersection of the two curves gives the equilibrium number of entrepreneurs for example all potential entrepreneurs with returns greater than or equal to $50,000 choose entrepreneurship. The reason for the downward sloping curve is because as there are lower returns to entrepreneurship, there are fewer entrepreneurs with at least that return.

This illustrates how extractive economic institutions affect overall entrepreneurship in the economy. This individuals returned entrepreneurship is $75,000, which was initially above the opportunity cost schedule. However, with insecure property rights, she can only make $25,000, which is less than the opportunity cost of $50,000. This makes entrepreneurship appear less attractive. There are several factors contributing to this reduction and returned entrepreneur ship, insecure property rights it's just one of them. Another one that we have already talked about is lack of protection for contractual agreements. Entrepreneurships rely on contracts to borrow money, and rely on the courts to uphold these business arrangements. Without these institutions, it is difficult to promote reliable contracts.

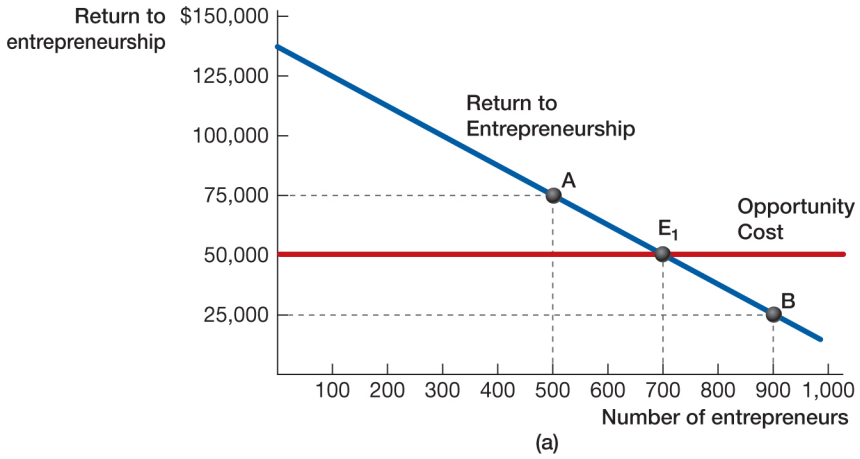

Figure 4.14: How Extractive Economic Institutions Reduce the Number of Entrepreneurs

Extractive economic institutions shift the return to entrepreneurship curve to the left. This is due to two potential reasons. One is weak property rights prevent entrepreneurs from capturing their full returns. Secondly, a lack of legal protection will prevent entrepreneurs from forming reliable contracts with business partners, which can reduce profitability by making supplies more expensive and revenues more unstable.

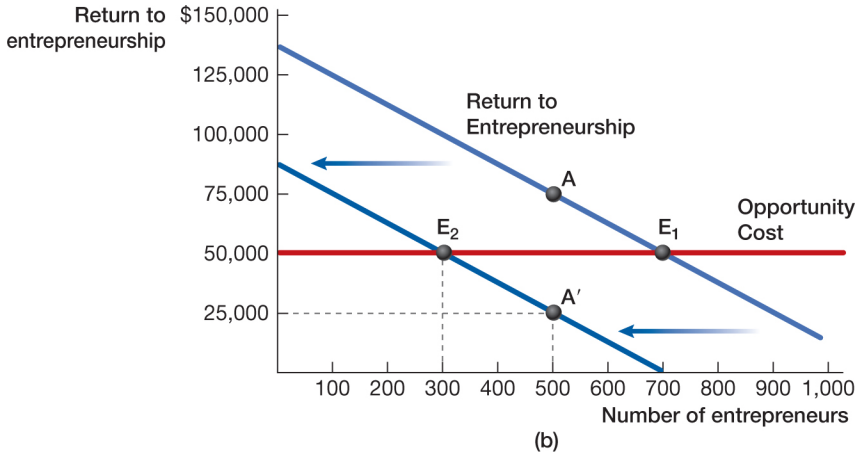

Figure 4.15: How Extractive Economic Institutions Reduce the Number of Entrepreneurs

Extractive economic institutions also shift the opportunity cost schedule upward because they introduced entry barriers that make entry into entrepreneurship more expensive. We can think of the cost of entry into entrepreneurship as the opportunity cost. As the opportunity cost increases, more entrepreneurs are discouraged from entering the market. This graph shows the overall impact of extractive economic institutions on the equilibrium number of entrepreneurs resulting from a leftward shift of the return to entrepreneurship curve and an upward shift of the opportunity cost schedule.

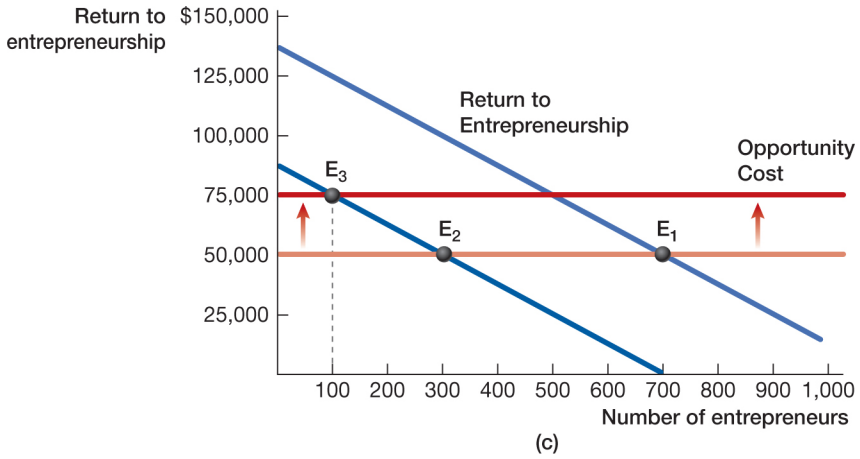

Figure 4.16: How Extractive Economic Institutions Reduce the Number of Entrepreneurs

In conclusion, extractive economic institutions discourage entrepreneur ship via two mechanisms:

1. By creating insecure property rights and limiting legal back up, they make entrepreneurship less profitable and shift the return to entrepreneurship schedule to the left.

2. By directing entry barriers, they make entry more costly and shift the opportunity cost schedule upward.

5 Week 5

5.1 Employment Unemployment

Measuring Employment and Unemployment

Unemployment can have significant economic and social costs, therefore it is very important to measure its presence in an economy. To do so, economists have agreed on a standard, though imperfect, method of defining employment and unemployment. In the United States, the standard is set by the Bureau of Labor Statistics in the Department of Labor, which tracks the official employment statistics for the US economy.

First, we need to define the population of potential workers. The population of potential workers includes the general population, but excludes three groups:

1. children under 16 years of age

2. people on active duty in the military

3. people who are living in institutions where the residents have restricted mobility (for example, facilities that provide long-term medical care or prisons)

The number of potential workers in the United States in January 2020 was 259.5 million people. This is also known as the civilian non-institutional population 16 years and older. Potential workers can be divided into three sub groups: employed workers, unemployed workers and those not in the labor force. People who have full-time or part-time pay jobs are considered to be employed workers. People who do not have a paid job, have actively looked for work in the prior four weeks, and are currently available for work are considered unemployed workers. Laid off workers are only considered unemployed if they are actively looking for a new job. Similarly, students and parents who don't have a paid job are only considered unemployed if they are actively looking for a job and are currently available to work, even part-time. In January 2020, there were 6.5 million unemployed workers in the United States. The labor force is the sum of all employed and unemployed workers.

Finally, all potential workers who do not fit the criteria for being unemployed or unemployed are classified as not in the labor force. People in this category of potential workers who do not have a paid job and are not looking for one, such as stay at home parents, disabled workers, many retirees and many students.

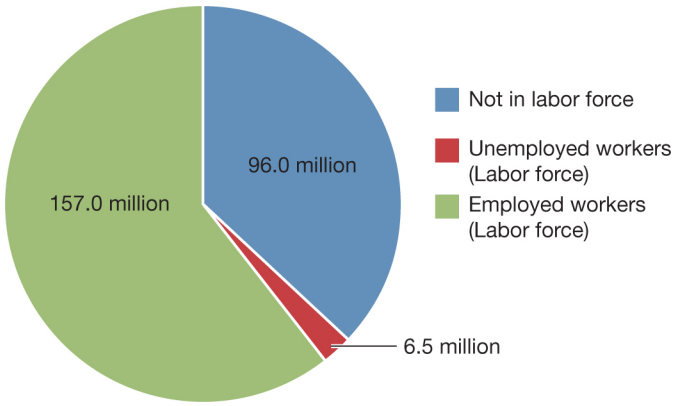

Figure 5.1: The Composition of the U.S. Population of Potential Workers (January 2020)

Using these definitions, we can compute the unemployment rate. The unemployment rate is defined as a percentage of the labor force that is unemployed.

$$UnemploymentRate = 100\% * \frac{Unemployed}{LaborForce}$$

$$UnemploymentRate = 100\% * \frac{6.5million}{163.5million}$$

$$UnemploymentRate = 4.0\%$$

Similarly, the labor force participation rate is defined as the percentage of potential workers who are in the labor force.

$$LaborForceParticipationRate = 100\% * \frac{LaborForce}{PotentialWorkers}$$

$$LaborForceParticipationRate = 100\% * \frac{163.5million}{259.5million}$$

$$LaborForceParticipationRate = 63\%$$

Although these two metrics are the most popular methods to measure the health of the labor force, they overlook two groups of important workers: discouraged workers and the underemployed.

Discouraged workers are potential workers who would like to have a job but have given up searching for one. Because they are not actively looking for work, these workers are not included in the unemployment rate. Instead, discouraged workers are counted officially as out of the labor force entirely. They were 350,000 discouraged workers in the United States in January 2020.

Similarly, we count all paid workers as employed, even if they would like to work more hours. Many workers in difficult economic situations would like to work more hours but do not have the option to do so. Although such workers are underemployed, they are not included in the official unemployment statistic. There were 4 million underemployed workers in the United States in January 2020.

The unemployment rate fluctuates with the economy. When the overall economy suffers from a recession, a period in which GDP falls, the unemployment rate tends to rise. During typical recessions, the unemployment rate reaches a level between 6% and 9%. When the economy is healthy and expanding, the unemployment rate eventually reaches a level between 3% and 5%. For example, the 2009 recession lead to a sharp rise in the unemployment rate and a peak rate of 10% in October 2009. During the Great Depression of the 1930s, the unemployment rate reached 25%. During the Covid recession, the unemployment rate rose to a peak of 14.7% in April. You can see the trends of the unemployment rate in the graph below.

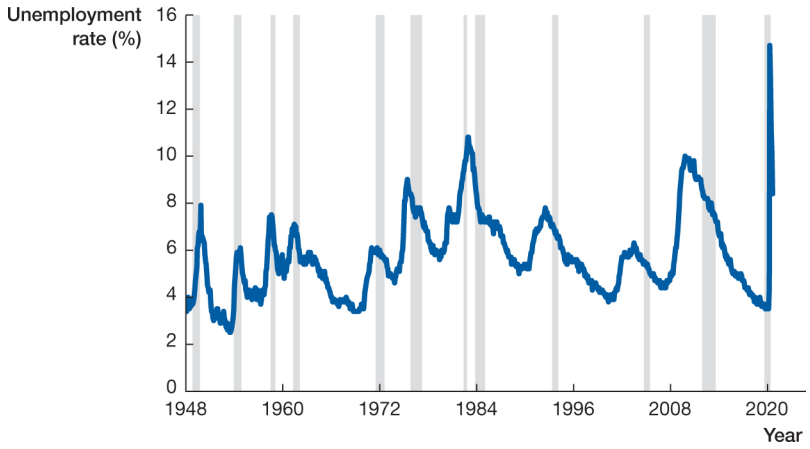

Figure 5.2: The U.S. Unemployment Rate Since 1948

It is also noteworthy to mention that the unemployment rate is never close to zero. Since 1948, the U.S. unemployment rate has gone below 3% during only one period, the early 1950s. Even during the economic boom in the 1990s, the unemployment rate reached a low value of 4%. It is a necessary attribute of a well functioning modern economy to have some unemployment.

Equilibrium in the Labor Market

Just like with any other market, we can analyze the dynamics of the labor market. In this market, the price of labor is the wage rate. Households supply labor, while firms demand labor. Firms are on the demand side because they need to hire workers for production. Optimizing firms try to maximize profits, therefore they demand the quantity of labor that produces the greatest feasible profit. Profit is calculated as revenues minus costs. How does a firm determine the profit maximizing quantity of labor? It compares the revenue that a worker produces with the cost of employing that worker.

Let's consider the example of a barber shop. If the shop has one barber, and that barber can produce $25 of revenue per hour from providing haircuts, this $25 is the value of the marginal product of labor of this worker. This represents the contribution of an additional worker to a firm's revenues. Let's assume the market wage for barbers is $15 per hour. Therefore, by employing this barber, the barbershop earns $10 in profit per hour. This is because $25 - $15 is $10 per hour.

If a barbershop considers hiring another barber, although this sounds like a good idea, there may not be enough customers to keep both barbers completely busy. For example, let's say the first barber has a fully booked out schedule but the second barber does not. Therefore, the second barber can only generate $20 in revenue per hour. The barbershop will gain $5 in profit per hour, because $20 - $15 = $5 per hour. So, it would still be a good decision to hire this second barber even if they are not as occupied as the first barber because the shop still makes a $5 profit.

What if the barbershop wants to hire a third barber? Although the third barber will increase sales a bit more, it will do so at a reduced margin compared to the second barber. Suppose that this third barber's value of marginal product of labor is $10 per hour. Because the market wage is

$15 per hour, then the barbershop loses money ($5) on the third hire. Therefore, it would be in the shop's best interest to not hire this third barber.

This is the idea of the diminishing marginal product of labor. This means that each additional worker creates less marginal output than the worker hired before them. For example, additional barbers will increase the number of haircuts that the barbershop can offer, but each additional barber will not be as productive as the one before it. If the barbershop faces a constant price for haircuts, the lower marginal output of additional workers also translates into diminishing value of marginal product of labor. Because the value of the marginal product of each additional barber is diminishing, hiring more barbers increases the total revenue of the barbershop by a reduced amount with each new hire.

Another important takeaway is that a firm hires workers up until the point that it cannot increase profits by hiring an additional worker. The firm keeps hiring as long as the revenue that an additional worker brings in for the firm is at least as great as the cost of employing that worker, which is equal to the market wage. We can see this in the graph below, which plots the value of the marginal product of labor against the number of workers employed. Because the value of the marginal product decreases as the number of workers employed increases, the curve is slopes downward.

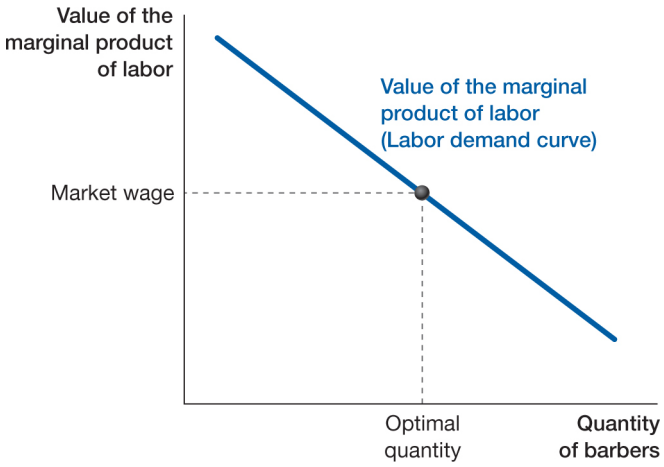

Figure 5.3: The Value of the Marginal Product of Labor is the Labor Demand Curve

The profit maximizing firm will hire the amount of labor at the point where the value of the marginal product of labor is either equal to or greater than the market equilibrium wage. As we change the market wage, the quantity of labor demanded moves along the curve depicting the value of the marginal product. The firm adjusts the number of workers and employers to make the value of the marginal product equal to the wage. Therefore, the downward sloping curve also represents the labor demand curve in addition to the value of the marginal product of labor because it shows how the quantity of labor demanded varies with the wage. A movement along the demand curve occurs when the wage changes and no other economic variables change other than the quantity of labor demanded.

Shifts in the Labor Demand Curve

There are a variety of things that can shift the labor demand curve entirely. Any change that affects the schedule relating the quantity of labor and the value of the marginal product of labor will shift the labor demand curve. We'll cover four factors that can shift the labor demand curve.

1. Changing output markets

   When the demand for haircuts shifts to the right, then the quantity of haircuts demanded at any given price increases, and therefore the price of haircuts increases. Accordingly, there is a rightward shift in the demand for haircuts which causes a rise in the value of the marginal product of barbers and a rightward shift in the demand curve for barbers.

2. Changing productivity

   When changes in productivity increase the value of the marginal product of labor, the labor demand curve shifts to the right. For example, if there are improvements in technology that can improve the productivity of barbers, this would shift the labor demand curve to the right. Technological progress can contribute to this.

3. Changing input markets

   When the cost of inputs goes down, businesses can purchase more of these inputs. This usually increases the marginal product of labor, which would shift the labor demand curve to the right. For example, if mechanical hair clippers suddenly became cheaper, barbers could purchase more of these items and potentially cut hair faster, which would improve their productivity. This would increase each barber's value of marginal product and shift the labor demand curve to the right.

The Supply of Labor

The labor supply curve represents the relationship between the quantity of labor supplied and the wage. It is derived from the principles of optimization. In this case, workers optimally allocate their limited time between paid work, leisure and other activities, which might include a home production such as childcare and cooking. If market wages are higher, it makes sense for workers to spend more time working outside the home. This kind of reasoning implies that as the wage increases, quantity of labor supplied increases. Therefore, the labor supply curve is upward sloping. It is possible for the labor supply curve to be completely vertical if the wage is high enough.

Figure 5.4: Upward Sloping Labor Supply Curve

In reality, the slope is much more smooth compared to this kinked version above. The vertical portion of the labor supply curve captures the fact that it becomes much harder to increase the quantity of labor supplied if almost all people who are interested in working have already found a full-time job. In other words, this is the idea that once wages reach a certain point, you can't use wages to attract more workers. In other words, wages have an implicit ceiling of appeal to increase your labor supplied. Is certainly true for certain occupations. For example, higher-paying occupations such as attorneys at law may not be affected by a small increase in salary. However, low-paying occupations such as service industries may attract more workers from an increase in wage.

Shifts in the Labor Supply Curve

A movement along the labor supply curve occurs when the wage changes and no other economic variables change. Any change that affects the entire schedule relating the quantity of labor supplied and the wage will shift the labor supply curve. We'll discuss three factors that can shift the labor supply curve.

1. Changing tastes

   Changing tastes or norms can affect people's willingness to take a paying job. For example, before World War II, working for pay outside the home was frowned upon if you were a married woman. However, this changed after the war.

2. Changing opportunity cost of time

   Innovations have changed the time needed for home production, such as vacuum cleaners. These can lower the opportunity cost of working outside the home by freeing up time that was previously needed for home production activities. This changes the opportunity cost of time and has been a factor contributing to the rise in female labor force participation.

3. Changing population

Increases in the size of the population which increase the number of potential workers in the economy, can also shift the labor supply curve to the right. One factor increase in population is immigration. For example, the United States experiences a net immigration inflow of roughly 1,000,000 people each year.

Putting Together Labor Supply and Demand

Equilibrium in a labor market is the point of intersection between the labor supply and labor demand curves. **At the competitive equilibrium wage, w*, the quantity of labor supplied is equal to the quantity of labor demanded.** In this case, all workers can work as many hours as they wish at this wage, and all firms are able to hire as many hours of labor as they need to. In contrast, at a wage above w*, the quantity of labor supplied would exceed the quantity of labor demanded and push the wage down. At a wage below w*, the quantity of labor demanded would exceed the quantity of labor supplied and push the wage up. This labor market equilibrium is also referred to as the equilibrium employment. We also referred to the equilibrium wage is the market clearing wage. This means that every worker who wants a job can find one.

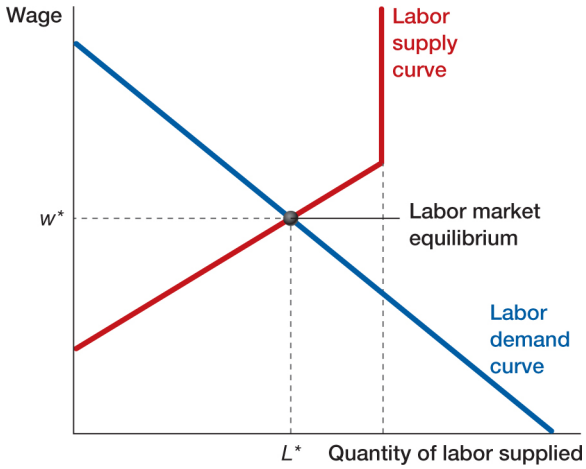

Figure 5.5: Competitive Equilibrium in the Labor Market

The market demonstrated in this graph is sometimes referred to as a "frictionless" labor market. In these kinds of markets, firms can instantly hire and fire workers, both workers and firms have complete information about each other and the wage adjusts instantly to clear the market.

Voluntary & Frictional Unemployment

Voluntary Unemployment

Official unemployment statistics are probably counting some workers who are voluntarily unemployed. They are willing to work, but only for a wage that is above the market clearing wage of w*. Therefore, at the equilibrium wage, they are happy to remain unemployed. Although this is

not common, it does occur.

Frictional Unemployment

Up until now we have assumed that the job market is frictionless. This implies that a worker could instantly find an employer if she is looking for work. However, this is not the case in reality. The job search process requires significant legwork in finding the employer that fits best for you. It takes time to find the right match, this contributes to friction in the job market. Frictional unemployment refers to unemployment that arises because workers have imperfect information about available jobs and need to engage in a time-consuming process and job search.

Wage Rigidity and Structural Unemployment

Structural unemployment occurs when the quantity of labor supplied exceeds the quantity of labor demanded. This can occur when the wage is fixed above the market clearing wage. This is an example of wage rigidity, where the wages are held fixed above the competitive equilibrium level. Good examples of this are minimum wage laws.

In the United States, we have federal national minimum wage laws. In January 2020, the federal minimum wage was $7.25 per hour. Each state can set their own minimum wage, so long as it is above this level. Washington has the highest minimum wage of $13.50 per hour.

Minimum wages will prevent the quantity of labor supplied to equal the quantity of labor demanded. For example, if the minimum wage is above the market clearing wage, then there will be a higher quantity of labor supplied than labor demanded. This means that if minimum wages are higher than the equilibrium wage, there will be far more people willing to work than there are positions open for hiring. Firms will not be able to meet this supply and therefore we'll still see unemployment occurring. These unemployed workers are willing to work at wages lower than the minimum wage, however employers are legally unable to higher them at these lower wages. Therefore, we see an unemployed population when minimum wages are higher than the market-clearing wage of w*. These workers are considered to be involuntarily unemployed. This is because the minimum wage laws are preventing them from obtaining full employment.

Minimum wage laws can have advantages and disadvantages. In the overall workforce in the U.S., including all education levels, only 1% of workers are paid the minimum wage. Therefore, the impact of the minimum wage on the labor market is modest. The minimum wage does prevent the market from reaching equilibrium for some types of low skilled workers but has little impact on the overall general labor market. We can see the effect of minimum wage laws on the labor market in the graph below. When the minimum wage is above the market clearing wage, the quantity of labor supplied exceeds the quantity of labor demanded, creating unemployment. The quantity of labor demanded is the point on the labor demand curve that intersects with a horizontal minimum wage line.

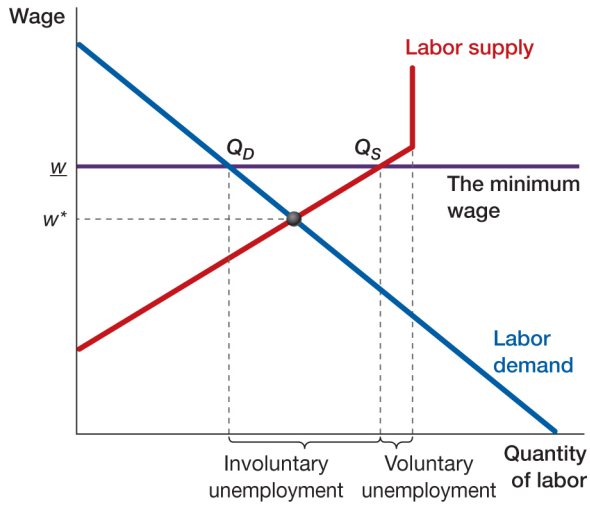

Figure 5.6: Labor Supply and Labor Demand in a Market with a Minimum Wage

Labor Unions and Collective Bargaining

Labor unions use collective bargaining to negotiate higher wages for their members. Collective bargaining has the same effect on unemployment as do minimum wage laws. If they can keep the negotiated wage above the market clearing equilibrium wage, the unions can cause the quantity of labor supplied to exceed the quantity of labor demanded, thus creating structural unemployment. In 2018, 10.5% of all employed workers in the United States are members of labor unions. In Italy, in comparison, 34.4% of all employed workers are members of labor unions.

Efficiency Wages

In 1914, the Ford motor company doubled its daily wage for most of their employees from $2.34 per hour to five dollars per hour. Why did they do this? They explained that they wanted the business to be on a lasting foundation, and that they were building for the future. This is an example of an efficiency wage. Efficiency wages are implemented by employers to increase motivation and productivity. This is a good guard against workers who may slack off on the job. A firm might be incentivized to pay higher wages to improve productivity in the workplace. Although a higher wage does not improve profits for the firm, it can increase productivity for the company in a variety of ways.

First, efficiency wages reduce worker turnover. Recruiting new workers is costly to the company. If workers are paid more than the equilibrium wage, they are motivated to keep their job because they would face lower wages elsewhere. Second, the fear of losing a high paying job motivates employees to work harder than they otherwise would, improving their productivity and increasing their overall hourly output. Third, some employees are grateful for a higher wage, leading them to reciprocate this generosity in the form of improved productivity in the workplace. Finally, higher wages improve the quality of the pool of workers who apply for the job in the first place. It's important to remember that all of these rest on the assumption that a higher wage would motivate an individual to work more productively. The main difference between efficiency wages

and minimum wages, although both are above the equilibrium wage, is that a firm voluntarily chooses to provide an efficiency wage. A firm does not voluntarily choose to provide a minimum wage, they are mandated to do so by the government.

Downward Wage Rigidity

Downward wage rigidity arises when workers resist a cut in their wage and firms respond to this resistance by holding wages fixed, and instead of firing workers. For example, if there is an outside force that shift the labor demand curve to the left, such as replacing workers with robots, this would reduce a firms demand for labor. Therefore, this SHOULD reduce the market equilibrium wage. However, in the case of downward wage rigidity, a firm may just fire workers instead of reducing their wages. We can see this demonstrated in the graph below. Instead of reducing the wage paid, the firm instead reduces the number of workers hired. Therefore, the quantity of labor demand it moves from L1 to L2.

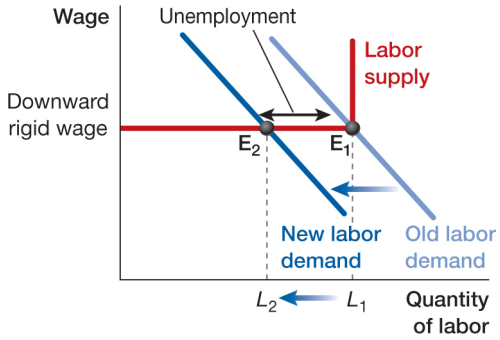

Figure 5.7: Shifts in Labor Demand Affect Equilibrium in the Labor Market

Cyclical Unemployment and the Natural Rate of Unemployment

The long run rate of unemployment is the average historical rate of unemployment. In the United States, the long run rate of unemployment hovers around 5.8%. The long run rate of unemployment includes frictional unemployment, as well as long-term structural unemployment. Long-term structural unemployment is considered to be economically inefficient. Structural unemployment is what we just talked about where the quantity of labor supplied exceeds the quantity of labor demanded. More people want to work than firms are able to hire.

This is different from the natural rate of unemployment, which is 4%. The natural rate of unemployment is the point at which a health economy hovers around. At 4% unemployment, an economy can still experience economic growth. The long run rate of unemployment is typically higher than the natural rate of unemployment.

When the current unemployment rate deviates from the long run rate of unemployment, this is cyclical unemployment. Cyclical unemployment rises in recessions and falls in economic booms.

The long run rate of unemployment should not be confused with the rate of unemployment that is socially optimal. This is because the long run rate of unemployment include some of inefficient sources of unemployment, such as structural unemployment. Structural unemployment is caused by things like downward wage rigidity. In these cases, firms would rather fire workers and reduce their wages. In these kinds of economies, there will be a relatively high level of structural unemployment and therefore a high long-term average rate of unemployment.