Economic Analysis III: Intermediate Macroeconomics

June 2021

Contents

1 Introduction

1.1 Growth Theory: A Historical Overview

Why do economies grow, and what explains why some economies grow faster than others? The study of growth can be traced all the way back to Adam Smith, who argued in 1776 that economies could grow faster if the constraints imposed on them by the state were relaxed. Smith argued that the division of labor and the increasing size of world markets would usher rapid economic progress, and emphasized the importance of increasing factor supplies in increasing quantities of output. In 1817, David Ricardo put forward the idea of diminishing returns to fixed factors - in his case, land - which would become the cornerstone of the Neoclassical growth model. He emphasized the importance of improvements in technology as the ultimate source of growth in this world. Ricardo's approach to measuring value and growth would later be the basis upon which Karl Marx would construct a theory of political economy and income distribution.

Modern growth theory probably began in 1928, when Frank Ramsay exposited and solved for a theory of "optimal" saving and growth. Keynes' exposition of a theory of output, employment and investment in 1936 was extended to a dynamic setting by Roy Harrod in 1939 and Evsey Domar in 1947. The Harrod-Domar model argued that growth could be sustained through the accumulation of capital, which would typically be in short supply relative to labor. A key idea in the Harrod-Domar model was that an increase in the saving rate, which translates into an increase in the investment rate, can raise the growth rate of the economy as a whole[1].

In 1956 and 1957, Robert Solow's celebrated model emphasized, once again, the role of diminishing returns. Solow argued that if capital was subject to diminishing returns, then changes in the saving rate in the short term would not affect long run growth rates - the only factor that could generate long-run growth was technical progress, which Solow's analysis took as given. Building on this insight, growth theory shifted focus to determining what factors could affect productivity, leading to Lucas's 1988 piece on human capital, Paul Romer's 1986 analysis focusing on the role of ideas and his 1990 contribution on the role of increasing returns to knowledge production, and Jones' 1995 contribution on eliminating scale effects in economic growth. Modern growth theory seeks to also understand the diffusion of ideas across countries, and focuses on the microeconomics of innovation, emphasizing the role of patent protection and monopoly rents in incentivizing R&D investment by firms.

[1] An important consideration in Harrod's analysis was the conflict between the notion of the "natural" rate of growth, which is the rate of growth consistent with technology and growth in factor supplies, and the "warranted" rate of growth associated with savings and investment. In Harrod's model, if the warranted rate of growth falls below the natural rate of growth, the economy suffers from growing capacity underutilization. In the other case, it suffers from rising unemployment.

1.2 Business Cycle Macroeconomics: A Historical Overview

Before the 1930s, there was no distinct field of Macroeconomics[2]. The question of what was responsible for the repeated fluctuations in aggregate economic activity was an unsolved question - according to classical theory, macroeconomic variables like employment and output were "supply-driven" and determined within a *general equilibrium* system. Friedrich Hayek, writing in 1933, emphasized the need to integrate the formidable mathematical apparatus underlying general equilibrium theory[3] with the study of what at the time was known as the "trade cycle."

After the Great Depression and the publication of Keynes' *General Theory*, it became clear that the analysis of the aggregate economy could not be adequately performed using the tools of classical economics. In particular, Keynes, his contemporaries and his followers emphasized the importance of *uncertainty* and the role of expectations in the determination of economic equilibrium, as well as of the role of fluctuations in "autonomous demand". Prior to Keynes, the attention of "trade cycle theorists" was devoted to understanding the *sources* of business cycles. After him, proponents of the Keynesian view attempted instead to understand how to "fix" business cycles, rather than to understand the sources of cycles themselves. This view, which became popular in policy cycles, held that understanding why business cycles happen was less important than understanding how to respond to them with policy, irrespective of the source of the cycle. According to Keynesians, capitalist economies had natural forces that could lead to instability or collapse absent the hand of a paternalistic state to steady their course.

Macroeconomic policymaking under Keynesian policymakers generally involved the construction of large-scale "macro-econometric" models with a large number of equations involving macroeconomic aggregates[4]. These equations were largely "behavioral" equations, in that they *assumed* relationships between macroeconomic aggregates with parameters that were fixed. While individual equations in these models sometimes derived from deeper motivating models, the motivating models for different equations were sometimes mutually inconsistent. However, through the 1960s, these inconsistencies were ignored, and the large-scale models were utilized to justify and implement activist fiscal and monetary policy geared towards stabilization of the cycle.

A key tenet of Macroeconomic policymaking in this period was the Phillips curve, a negative relationship between inflation and unemployment which was thought to represent a policy trade-off facing any government. In the 1970s, the US underwent a period of high inflation accompanied with high unemployment, which violated this key principle. At the same time, now influential critiques[5] of the large scale macroeconometric models used in policymaking appeared[6]. These critiques led to

[2]For more details, see Lucas (1977).

[3]General Equilibrium theory was largely developed by the members of the Lausanne School, based at the University of Lausanne, and followed the intellectual achievements of Lèon Walras and Vilfredo Pareto. See here, for instance. Further important contributions to General Equilibrium theory were made by Kenneth Arrow and Gerard Debreu.

[4]See Beatrice Cherrier's piece here, for some examples of what this effort looked like in practice.

[5]See, for instance, Lucas and Sargent (1979).

[6]The most important of these was the Lucas Critique, which held that one could only use parameter

a shift in the way Macroeconomics was done along multiple dimensions, and are largely responsible for the state of Macroeconomics today, which is a continuation of the so-called "Neoclassical Synthesis".

Macroeconomic models today are

- **Microfounded**, with all aggregate variables determined by aggregating over optimizing choices made by agents throughout the economy.
- **Dynamic**, with an explicit role for expectations of the future and anticipation of future changes feeding into decisions made by agents today.
- **Stochastic**, with an explicit role for various shocks to the economy coming from unanticipated changes in variables.
- **General Equilibrium**, requiring all aggregate variables, particularly prices, to satisfy internal consistency conditions which guaranteed simultaneous equilibrium in all markets under study.

While the *methodology* underlying Macroeconomics is now largely a matter of consensus, the set of topics the field studies has changed several times, leading to changes in the structure of major models. In the 1990s and early 2000s, economies around the world experienced robust growth, low unemployment and inflation, a period today known as the Great Moderation. The Great Recession in 2007-08 led to a deeper appreciation of the special role played by the financial system, which now features prominently in models of the business cycle. The COVID-19 crisis led to a renewed focus on the supply side of the economy and how it interacts with the world economy. As global inequality rises, macroeconomists have begun trying to understand how inequality and policy interact with each other. As further crises affect the world, newer models will doubtless evolve to try to understand them.

1.3 Integrating Growth and Business Cycle Theory

Econ 52 provides an integrated view of Growth and Business Cycles. We do this by starting with the workhorse model in Macroeconomics, the Neoclassical Growth Model. We use this simple model to study the effects of fiscal policy and in the study of long-run economic growth. We then consider the Real Business Cycle (RBC) model, which augments the Neoclassical Growth Model by adding shocks and endogenizing labor supply realistically. We conclude by introducing sticky-price extensions of the RBC Model. Further classes at Stanford extend the RBC model to a multi-country setting, and graduate school extends these models to economies with realistic heterogeneity across households.

estimates obtained by econometric techniques if the equations being estimated were the outcome of an underlying model and the parameters being estimated were related to the primitives of the model, such as preferences or technology.

1.4 Mathematical Preliminaries

We start with some math preliminaries that will be required throughout the course. You will need some prior knowledge of calculus to understand them. This is the least exciting of all the material you will see in the course, but it is absolutely necessary to get the maximum out of the rest. Please contact your TA(s) if you have any issues with the math described here.

1.4.1 Proof Strategies

Like any other mathematical subject, a large part of macroeconomics consists in trying to assess the truth value of *conditional* mathematical statements, which are statements of the form "if p then q". In this example, p is sometimes called a "hypothesis" and q is sometimes called the "conclusion." For example, we might be interested in proving the claim

> Let $x > 1$ be an integer. If x is an even integer, then x^2 is an even integer.

How can we *prove* such a statement? Here are the three most common ways.

- **Direct Proof:** Find a sequence of statements $r_1, r_2, \ldots, r_n$ such that p implies r_1, which implies r_2, and so on, until finally r_n implies q. Note that to prove p implies r_1, you must show that r_1 is a clear conclusion that follows from p. In the example above, a direct proof might look like this.

  $$\begin{aligned} & x \text{ is even.} \\ \implies & x = 2z \text{ for some integer } z. \\ \implies & x^2 = 2 \times 2z^2 = 2k \text{ for an integer } k. \\ \implies & x^2 \text{ is even.} \end{aligned}$$

  Note that in this proof, we use the result that the integers are closed under multiplication, which we haven't proved. In Econ 52, you can take as given mathematical statements that are "obvious" - we are more interested in the economics involved in your logic than the precise mathematical details.

- **By Contradiction:** A proof by contradiction assumes that the hypothesis is true and the conclusion is false, and then tries to identify a logical inconsistency. resulting from this. For our statement above, a proof by contradiction is the following.

  $$\begin{aligned} & \text{Suppose } x \text{ is even and } x^2 \text{ is not even.} \\ \implies & x = 2z \text{ for some integer } z, \text{ so } x^2 = 2 \times 2z^2. \end{aligned}$$

  This implies that 2 is a divisor of x^2, so x^2 is even. This contradicts the initial assumption that x^2 is not even.

- **Proving the Contrapositive:** It can be shown that if $p \implies q$, then $\neg q \implies \neg p$, where $\neg q$, read "Not q", is the *negation* of the statement q. The latter statement is called the *contrapositive* of the original statement. In some cases, proving a contrapositive can be much easier than a direct proof. In our example above, start by noting that the contrapositive of the statement is

  > Let $x > 1$ be an integer. If x^2 is not an even integer, then x is not an even integer.

  A contrapositive proof might look something like this.

  $x > 1$ is an integer such that x^2 is not even.

  $\implies$ Since x is an integer, it can be written as a product of prime numbers[7] $x = p_1^{m_1} p_2^{m_2} p_3^{m_3} \dots p_k^{m_k}$.

  $\implies$ Clearly, $x^2 = p_1^{2m_1} p_2^{2m_2} p_3^{2m_3} \dots p_k^{2m_k}$.

  $\implies$ Since x^2 is not even, none of the primes $p_1, \dots, p_k$ can equal 2.

  $\implies$ 2 is not a prime factor of x, so x is not even.

How can we *disprove* statements? This is (on the surface!) much easier: you can just find a contradiction, or find a *single* counterexample. Consider the statement "for all integers x, $x^2 < 1000000x$." Two pretty simple "dis-proofs":

- **Disproving by contradiction:** We have,

  $$x^2 < 1000000x \implies x(1000000 - x) > 0 \implies -1000000 < x < 1000000$$

  which clearly excludes all integers z such that $z > 1000000$ or $z < -1000000$. Since this latter set is nonempty, we have a contradiction to the hypothesis that the statement is true for all integers.

- **Disproving by Counterexample:** The statement is not true for $x = 1000001$, which is indeed an integer.

A common mistake when trying to prove a statement is to examine a couple of cases and, finding that the statements hold for those particular cases, to conclude that the statement is always true. This is incorrect. Again consider the statement "$x^2 < 1000000x$". This statement isn't correct in general, but it does hold for every $x \in (-1000000, 100000)$ - so a naive check could take a while to find a counterexample, if it ever did!

1.4.2 Some Useful Derivatives

$$\frac{d}{dx}x^n = nx^{n-1} \qquad \frac{d}{dx}e^{f(x)} = f'(x)e^{f(x)}$$

[7] The fact that any positive integer greater than 1 can be uniquely factored into primes is a result called the *Fundamental Theorem of Arithmetic*.

$$\frac{d}{dx}ln(x) = \frac{1}{x} \qquad \frac{d}{dx}a^x = (lna)a^x$$

Let $f(x)$ and $g(x)$ be two functions:

$$\frac{d}{dx}(f+g) = f' + g' \qquad \frac{d}{dx}(f.g) = f'.g + f.g'$$

$$\frac{d}{dx}(f-g) = f' - g' \qquad \frac{d}{dx}(\frac{f}{g}) = \frac{f'g - fg'}{g^2}$$

1.4.3 The Chain Rule

Let $h(x) = f(g(x))$. Then,

$$\frac{d}{dx}h(x) = \frac{d}{dx}f(g(x)) = f'(g(x)) \cdot g'(x)$$

Some Examples: (i) $f(x) = ln(x)$ and $g(x) = x^2$. Then, $h(x) = ln(x^2)$.

$$\frac{d}{dx}h(x) = \frac{d}{dx}ln(x^2) = \frac{1}{x^2}2x = \frac{2}{x}.$$

(ii) $f(x) = (1-x)^2$ and $g(x) = (ln(x))^2$. Let $h(x) = f(x) \cdot g(x)$.

$$\frac{d}{dx}h(x) = f'g + fg'.$$

By the chain rule, $\frac{d}{dx}f(x) = -2(1-x)$. [You can think of the function f as the composition of two functions: $\tau(x) = (1-x)$ and $\beta(x) = x^2$. Then, $f(x) = \beta(\tau(x))$, and the derivative follows from the chain rule.]

Similarly, by the chain rule, $\frac{d}{dx}g(x) = \frac{2ln(x)}{x}$ ($\tau(x) = ln(x)$, then $g = \beta(\tau(x))$, and the derivative follows from the chain rule).

1.4.4 Logs, Elasticities and Growth Rates

We will repeatedly make use of logarithms in Econ 52, particularly owing to how easy they make dealing with two objects of particular interest to macroeconomists: elasticities and growth rates. Here's a quick refresher.

- The **natural logarithm**[8] of x, denoted by $\log x$, is the number z satisfying $x = e^z$ where e, sometimes called Euler's number, is a positive real number equal to approximately 2.718. Logarithms satisfy three particularly useful properties:

[8] In Econ 52, unless otherwise stated, the notation $\log x$ will always refer to the natural logarithm of x, and not the logarithm to the base 10. If this confuses you, just replace the notation $\log x$ with $\ln x$ everywhere in these notes.

- The Product/Quotient Rule: $\log(xy) = \log x + \log y$
- The Exponentiation Rule: $\log(x^a) = a \log x$
- The Approximation Rule: $\log(1+x) \approx x$ for x close to 0.

The Approximation Rule can be proved using a first-order Taylor expansion of the function $f(x) = \log(1+x)$ around $x = 0$. To see this, note that by Taylor's theorem, for x close to 0,

$$\begin{aligned} f(x) &\approx f(0) + x f'(0) \\ &= \log(1+0) + x\frac{1}{1+0} \\ &= x \end{aligned}$$

- The **Elasticity** of variable Y with respect to X is the quantity

$$\varepsilon_{Y,X} = \frac{X}{Y}\frac{dY}{dX} = \frac{dY/Y}{dX/X} = \frac{d\log Y}{d\log X}$$

which is a property that will be useful.

- The **Growth Rate** of variable Y is

$$g_t^Y = \frac{Y_{t+1} - Y_t}{Y_t}$$

Observe that

$$\begin{aligned} 1 + g_t^Y &= \frac{Y_{t+1}}{Y_t} \\ \implies \log(1 + g_t^Y) &= \log Y_{t+1} - \log Y_t \\ \implies \log Y_{t+1} - \log Y_t &\approx g_t^Y \end{aligned}$$

where the last line applies the Approximation Rule, presuming that g_t^Y is relatively small. This will prove particularly useful when we try to find growth rates of complicated looking things. For example, suppose we know that the variables X_t, Y_t, Z_t are growing at 2%, 3% and 1% respectively, and we want to find out how fast the variable $W_t = \frac{\sqrt{X_t} Y_t^3}{Z_t^5}$ is growing. While you're welcome to try it out directly, here's the easiest way: note that

$$\log W_t = \frac{\log X_t}{2} + 3\log Y_t - 5\log Z_t$$

Subtracting the date-t version of this equation from the date-$t+1$ version,

$$\begin{aligned} \log W_{t+1} - \log W_t &= \frac{\log X_{t+1} - \log X_t}{2} + 3(\log Y_{t+1} - \log Y_t) - 5(\log Z_{t+1} - \log Z_t) \\ &\implies g_t^W = \frac{g_t^X}{2} + 3g_t^Y - 5g_t^Z = 1\% + 9\% - 5\% = 5\% \end{aligned}$$

It turns out that the formula for growth rates $\log Y_{t+1} - \log Y_t \approx g_t^Y$ is an exact formula if we treat time as continuous and growth as exponential. To see this, suppose Y_t grows exponentially at the rate g_Y, so that $Y_t = Y_0 e^{g^Y t}$. Dividing the date $t+1$ version of this equation by the date t version, we get

$$\frac{Y_{t+1}}{Y_t} = \frac{Y_0 e^{g^Y(t+1)}}{Y_0 e^{g^Y t}} = e^{g^Y} \implies \log Y_{t+1} - \log Y_t = g^Y$$

1.4.5 Optimization with Equality Constraints

You should be familiar with this material (from Math 51 or Econ 50). This is meant to refresh your memory or be a handy reference. Consider the following problem.

$$\max_{x,y} \sqrt{xy} - x^2 \text{ s.t. } y = 10$$

How would you solve this problem? A very simple way to do this is to substitute $y = 10$ into the objective function and maximize as a function of x only.

But what about

$$\max_{x,y} \sqrt{xy} \text{ s.t. } \frac{x+y}{\sqrt{x}+y^2} = 10$$

There is no way to reduce this into a problem of a single variable or to express one variable as a function of the other variable in order to get rid of the constraint. What we need here is a very clever tool called the "Lagrangian Function". In general, if the problem is

$$\max_{x,y} f(x,y) \text{ s.t. } g(x,y) = c$$

where f and g and are functions of x and y and c is a constant, then the Lagrangian is

$$\mathcal{L}(x,y,\lambda) = f(x,y) + \lambda(c - g(x,y))$$

where $\lambda \neq 0$ is called the Lagrange Multiplier. So, how does this help us? In math, a very commonly used technique is to reduce a problem we don't know how to solve to a form which we can handle. We all know how to solve the problem

$$\max_{x,y,\lambda} \mathcal{L}(x,y,\lambda)$$

It turns out that the (x^*, y^*) values that solve this new problem are the solution to the original problem as well! (In the new problem, you will also get an optimal value for λ). The proof of this claim goes beyond the scope of this course. The interested reader, however, can ask for a reference.

To solve the new problem, we simply look at the three first-order conditions[9]

$$\mathcal{L}_x = 0$$
$$\mathcal{L}_y = 0$$
$$\mathcal{L}_\lambda = 0$$

to get (x^*, y^*, λ^*). If you find more than one set of solutions to the above equations, then the one that gives you the highest value of f is the one you want (i.e., it will satisfy the second-order condition).[10]

The Lagrange Multiplier, λ, has a very useful economic interpretation. Let (x^*, y^*, λ^*) be the solution to the maximization problem. Then one can show that

$$\frac{\partial f(x^*, y^*)}{\partial c} = \lambda^*$$

Therefore λ^* tells you by how much your objective will increase if you add one more unit to the constraint. If you interpret f to be the utility function and $g(x, y) = c$ to be your budget constraint (c is your income with which you can buy goods x and y), then the Lagrange Multiplier tells you how much more utility you can get by having one more unit of income to spend. This is why λ is often called the *shadow price* of the resource: it's the utility cost of not having one more unit of income to buy more goods.

What about the second-order conditions? Well, they do exist but we shall not bother with them in this course (because you will always get problems for which the solution to the first-order conditions and constraints uniquely satisfies the second-order condition).

Now try solving the following problem using the method just described (even though there is an easier way to solve it):

$$\max_{x,y} \sqrt{xy} \text{ s.t. } x + y = 10$$

What if you had to find the minimum? i.e. How would you solve the problem

$$\min_{x,y} f(x, y) \text{ s.t. } g(x, y) = c$$

The answer is rather simple. Use the same method! Ideally, we should check the second-order conditions to make sure that we have a minimum or maximum, but that makes the math too complicated for this class. We'll concentrate more on the Economics and less on algebra skills.

[9] $\mathcal{L}_x$ is the partial derivative of $\mathcal{L}$ w.r.t. x, i.e. $\frac{\partial \mathcal{L}}{\partial x}$

[10] Conversely, if you were solving a minimization problem, the solution that gives you the lowest value of f is the one you want.

1.4.6 Infinite Geometric Sequences

In Econ 52, we will sometimes need to calculate the sums of infinite sequences of the form

$$S = 1 + \beta + \beta^2 + \beta^3 + \dots$$

where $\beta < 1$. There is a beautiful mathematical trick to make calculating these sums easy. Start by writing

$$\begin{aligned} S &= 1 + \beta + \beta^2 + \beta^3 + \dots \\ \beta S &= 0 + \beta + \beta^2 + \beta^3 + \dots \end{aligned}$$

Subtract the second line from the first, to get

$$(1-\beta)S = 1 \implies S = \frac{1}{1-\beta}$$

For bonus points: what goes wrong if $\beta \geq 1$?

1.5 Conceptual Issues

Before we begin discussing Macroeconomics, we need to cover a couple of key conceptual issues. For more detail on these, an excellent resource is Parker (2010).

1.5.1 Models, Exogenous Variables, Endogenous Variables and Parameters

A lot of the criticism surrounding Economics in general, and macroeconomics in particular, focuses on the use of models and the absence of "realism" in these models. To debate these issues sensibly, we first need to define a few concepts a bit more clearly.

A **Macroeconomic Model** consists of three sets of objects.

- A set of **Exogenous** variables, whose values the model takes as given.
- A set of **Parameters**, which are just exogenous variables whose values aren't expected to change a lot. Formally, there is no difference between parameters and exogenous variables: both describe quantities we take as given from the perspective of what we are trying to explain. Informally, parameters are variables whose values we expect to remain fixed both inside the model and outside the model as well, whereas exogenous variables have values we expect to change.
- A set of **Endogenous** variables, whose values can be calculated using the model, given values of parameters and exogenous variables.

- A set of equations that relate exogenous variables, endogenous variables and parameters together. The **solution** of a model is the set of values of endogenous variables that satisfies these equations, given values for the parameters and exogenous variables.

Exogenous variables and Parameters, together with any functional form assumptions we make, are also sometimes called the **primitives** of a model. The idea of primitives is that they summarize all assumptions and all values you take as given before you solve a model - once you have the model's primitives, you should be able to solve for all endogenous variables by hand (or using a computer). For example, in much of this class, we've assumed that production is Cobb-Douglas. This is a primitive of our model - we're not going to try to explain why production is Cobb-Douglas. A lot of discussions around papers centers on whether a model's primitives are sensible or not.

The **solution** of a model is an expression for each endogenous variable in terms of the exogenous variables and parameters of the model *only*. A solution is said to be in *closed form* if each of these expressions for the endogenous variables can be represented as an algebraic expression.

To fix these ideas, let's consider the simplest model in economics: demand and supply for a single good in a perfectly competitive market. Let p, y, S, D denote the price of the good, the consumer's income, the supply of the good at price p and the demand for the good at price p respectively. We specify the equations of the demand and supply curves as[11]

$$\begin{aligned} S &= cp \\ D &= ay - bp \\ S &= D \end{aligned}$$

In this simple model,

- We take the consumer's income y as exogenous. That is, we'll be taking income as given, and not try to explain where her income comes from.
- The model's parameters are the constants a, b, c. We don't try to explain where these come from. In a deeper model that starts from consumer preferences and optimization, a, b would be functions of the utility function's parameters, for example. But this is a simple model designed to study how price and quantity are determined *given these demand and supply functions*.
- The endogenous variables are prices and quantities, p, S, D.

[11]The first two equations would sometimes be called "behavioral relationships": they summarize the optimizing behavior of two agents, a firm and a household respectively, and are true both in equilibrium and off-equilibrium. Note that a modern macroeconomic model would need to contain a specification of the utility function and production function that produced these demand and supply curves as well. The third equation is an "equilibrium condition" - it must hold at $p = p^*$, but will not in general hold outside of it.

What do we do with this model? We solve its equations to get

$$p = \frac{ay}{b+c} \qquad ; \qquad S = D = \frac{ac}{b+c}y$$

which expresses our endogenous variables as functions of only exogenous variables and parameters.

What insight does this model now give us? Well, suppose the consumer gets richer, so y goes up[12]. The model tells us we should expect both prices and quantities to go up.

We can now look deeper at our model to understand, intuitively, what's going on. Looking at the supply curve equation, we see that there is no change at all - but looking at the demand curve equation, it's clear that quantity demanded at each price has risen, implying a rightward shift in demand. This should raise both prices and quantities. Thus, a higher income level raises prices and quantities by raising demand at each price level.

Naturally, the models we'll look at in Econ 52 are a bit more complicated than this. But the best way to understand a model is to always reduce it to its primitives and see how these primitives map into endogenous variables.

1.5.2 State Variables

When models are dynamic, a solution of a model typically involves equations that relate quantities over time. This leads to the concepts of state variables, which summarize the state of the economy at any point in time, and control variables, which represent choices made. A **state variable** is any variable, whether endogenous or exogenous, whose value is taken as given when solving for a set of dynamic equations at a particular point in time. A **control variable** is any endogenous variable whose values at date t are chosen at date t. Identifying state variables is something of an art, but a general rule of thumb is the following: state variables are "pre-determined", in that their values are either fixed, or determined *prior* to the current period. A common example is the capital stock in most dynamic macroeconomic models: it is common to assume that it takes time to build factories and install equipment, and thus the capital stock available for production at date t is determined by investments made at date $t-1$.

1.5.3 Partial vs General Equilibrium

Macroeconomics is all about General Equilibrium - that is, we look for a set of prices that simultaneously leads to equilibrium in all markets we are studying. The distinction between partial and general equilibrium is somewhat fuzzy in the literature, but in general,

- A model is a **partial equilibrium** model, if at least one of its variables or parameters is an object that should be determined in a separate market that is not

[12] Macroeconomists might call this a "positive income shock."

described in the model.

- A model is a **general equilibrium** model, if the variables it takes as given are not determined as equilibrium objects in markets not described in the model.

Partial equilibrium analysis typically abstracts from feedback effects that can arise when changes in one market lead to changes in other markets that can then feed back into changes in the original market. For a fantastic example of how general equilibrium analysis can completely overturn the result of a partial equilibrium analysis, see here.

2 Macroeconomic Data

This section introduces the key objects and concepts that macroeconomics studies and how these can be measured in the data. It introduces the concepts of output, employment and prices that are the key focus of macroeconomic analysis, and their empirical analogues.

2.1 Measuring Output

Most Macroeconomic data come from the National Income and Product Accounts (NIPA), a database maintained by the Bureau of Economic Analysis in the US. The NIPA provide a robust conceptual framework for the measurement of output and income, even in modern economies with hundreds of industries connected by complex supply chains, thousands of firms, and millions of workers.

2.1.1 GDP: A Definition

The key concept in the measurement of output is the **Gross Domestic Product** (GDP), defined as the **market value of all newly produced final goods and services produced within a given period of time by factors of production located inside a country.** GDP summarizes total production based on the **location** of the factors of production, not their **national origin**. For instance, if an Indonesian national working at a US firm in Palo Alto produces $1000 worth of output in a given day, this $1000 is included in the calculation of US GDP for the year. Note that:

- **GDP does not include the value of non-market activities.** This can make the interpretation of GDP challenging in economies with extensive home production or in ones with extensive informal markets.
- **GDP does not capture the value of transactions associated with the resale of already produced goods.** For instance, when an individual buys a used car from a friend, the transaction does not contribute anything to GDP[13].
- **GDP does not capture transactions associated with intermediate goods.** For instance, consider a supply chain where a car manufacturer purchases steel from a steel producer and uses it to make cars, which are sold to a final user. The value of the final cars sold are included in GDP, being a final good, but the value of the steel purchased by the manufacturer to make the car is not.
- **GDP does not capture transactions made by domestic residents abroad.** For instance, the value of any goods purchased by a US citizen on holiday in Mexico contributes 0 to US GDP.

[13] Note that this does not mean the used car industry contributes nothing to GDP, since the firms engaging in used car sales make profits on the sales of used cars which are then used to compensate workers working at these firms and owners and shareholders of the firms themselves. It is only the value of the cars themselves that is not included in GDP. See example on intermediate goods below.

2.1.2 Calculation of GDP: The Product/Value Added Method

The product method follows from the definition of GDP directly. To calculate GDP by the product method, take each transaction in the economy involving the production of new output[14]. Calculate the Value Added by the transaction *i* at date *t*, using

$$ValueAdded_{it} = Output_{it} - ValueIntermediates_{it}$$

The sum of value added by all production transactions performed within the US domestic territory equals GDP.

$$GDP_{Vt} = \sum_{i \in US} ValueAdded_{it}$$

Here's a simple example for how this works. Consider an incredibly simple economy inhabited by an iron ore company, a steel producer, a car manufacturer and a final consumer. The iron ore company extracts iron ore from the earth directly and sells \$20 worth of iron ore to the steel producer. The steel producer uses the \$20 worth of iron ore to produce \$50 worth of steel, which it sells to the car manufacturer. The car manufacturer sells a car worth \$100 to the final consumer.

How do we calculate GDP in this economy? Let's go through each transaction.

- Raw materials in the earth clearly don't have any value until they are mined. Thus, the value added by the iron ore company is the entire value of their output, \$20.
- The steel producer produces \$50 worth of steel, but buys intermediate goods - iron ore - worth \$20. Thus, its value added is \$50-\$20 = \$30.
- The car manufacturer sells output (cars) worth \$100 to the final consumer, but buys intermediate inputs (steel) worth \$50 from the steel manufacturer. Clearly, its value added is \$100-\$50=\$50.
- Total value added in the economy must therefore be the sum of value added by the three firms, which is \$20+\$30+\$50=\$100.

Observe that in this example, the value of GDP obtained by summing value added across all firms equals the value of sales by the final firm. This is not an accident - it is identically true that the sum of value added across all firms within the economy must equal the value of sales of final goods produced within the economy. The reason why the value added approach is preferable is because in practice, many firms sell the same product for use both as a final good by some buyers and an intermediate good by some others. For instance, a computer maker may sell computers directly to households - a transaction involving sales of final goods - but also sell computers to other companies who produce other products. The latter transaction would involve the sale of an

[14] As we noted, this excludes resale transactions that do not generate additional incomes for any agents.

intermediate good. However, it is hard to require firms to provide data on their sales by customer, and so typically the only data we have is on the purchases of inputs by the firm and sales by the firm. In this case, using the value added approach correctly accounts for the multiple uses of computers and only includes sales of computers as final goods.

Since GDP_{Vt} relies on data provided at the firm and establishment level on revenues and value of input purchases, it is usually only available with a lag.

2.1.3 Calculation of GDP: The Expenditure Method

The expenditure method for calculating GDP relies on the identity that the value of domestic sales of final goods and services must equal the sum of the value of total demand for these domestically produced goods and services, sometimes called Aggregate Demand. We define four key components Aggregate Demand.

- Consumption[15] C is the value of all consumer goods and services purchased by households to fulfill their immediate wants. It accounts for about 69% of US GDP, and is particularly stable over the business cycle.
- Investment[16] I is the value of spending by households and firms on new capital goods, which includes new structures (real estate, plants and office buildings), equipment (machinery and tools, software, etc) and on the accumulation of inventories[17]. Investment is only about 15% of US GDP, but is extremely cyclical.
- Government Consumption and Gross Government Investment G. In the US, government spending is about 19.5% of GDP, a somewhat low number among all OECD economies.
- Net exports, $X - M$, the difference between foreign demand for domestically produced goods and domestic demand for foreign produced goods. Exports and Imports in the US are about 13.5% and 17% of GDP respectively.

We can calculate the Expenditure-based GDP measure using

$$GDP_{Et} = C_t + I_t + G_t + (X_t - M_t)$$

GDP_{Et} is the most timely measure of GDP, released quarterly by the BEA.

[15] More formally, Personal Consumption Expenditure.

[16] More formally, Gross Private Domestic Investment.

[17] Firms are said to accumulate inventory when their sales of output within a given period are lower than the value of output produced in that period. In this case, the NIPA treats inventories as though firms are purchasing these unsold goods from themselves, and classifies these "purchases" as a form of investment since firms can raise sales by running these inventories down in the future.

2.1.4 Calculation of GDP: The Income Method

The Income Method relies on the identity that any transaction involves the generation of income for at least one of the factors of production in an economy. Recall that GDP is the sum of value added by each transaction, and that the value added of the transaction is the difference between the value of output and the value of intermediate inputs used up. Where does this added value go? It shows up as income for the individuals that produced the final good being transacted.

In Econ 52, we will consider economies with two factors of production, labor and capital. Income to labor accrues in the form of wages and salaries, while income to capital will accrue in the form of rental income for individuals owning the capital stock. In reality, the income approach to calculating GDP considers eight different kinds of income, which are:

- **Compensation of Employees**, which includes the value of wages and salaries and the value of contributions made by employers to social security benefits.
- **Proprietors' Income**, which is the income earned by owners of unincorporated businesses.
- **Rental Income**, which is income received by property owners.
- **Corporate Profits**, which is the retained earnings of corporations after all costs are accounted for.
- **Net Interest**, interest income to lenders to the corporate sector[18].
- **Indirect Taxes minus Subsidies**, which are income earned by the Government[19]
- **Net Business Transfer Payments**, which are transfers paid by businesses to others.
- **Net Surplus of Government Enterprises**, the excess of the value of publicly produced goods and services over their cost of production. This is typically negative.

The inputs for calculating GDP by the income method are available regularly via payroll and business tax data, which means that GDP_{It} is calculated on a quarterly basis.

[18] Note that net interest does not include consumer or government debt, since it is not assumed to flow from the production of new goods and services.

[19] Direct taxes are already counted in the other forms of income being taxed, since these are typically measured pre-tax.

2.2 Important Issues in the Measurement of GDP

2.2.1 Nominal vs Real Quantities

The distinction between nominal and real variables is important in order to interpret the value of any variable. A variable is said to be **nominal** if it is measured in terms of current prices, and is said to be **real** if it is measured in terms of a constant level of prices. Changes in a real variable reflect changes in quantities, whereas changes in nominal variables can conflate changes in quantities and changes in the prices associated with that variable.

To make this concrete, suppose the economy has one good, of which Y units are produced and each of which is sold at a price P. Real GDP in this economy is just Y, the number of units of the good produced. Nominal GDP is the dollar value of these Y goods, and equals PY.

How do we calculate real GDP in complex economies? The typical approach involves starting with Nominal GDP, which is readily available, and *deflating* it using a price index called the GDP Deflator. Given nominal GDP PY and the deflator P, we can calculate real GDP $Y = PY/P$. In practice, the construction of the GDP deflator is an involved process, and choosing the correct deflator can pose interesting conceptual issues (See Hard Questions for an example).

2.2.2 GDP vs GNP

GDP measures the value of domestically produced final goods. However, this is only equal to the value of income generated earned by domestically located factors of production. In particular, GDP excludes income earned by a country's citizens abroad, and includes incomes earned by foreigners within a country's borders. A better measure of national income is the Gross National Product (GNP), defined as

$$\begin{aligned}\text{GNP} &= \text{GDP} + \text{Net Factor Payments (NFP)} \\ &= \text{GDP} + \text{Net Capital Payments} + \text{Net Labor Payments} \\ &= \text{GDP} \\ &+ \text{Income from capital owned abroad} - \text{Income earned by foreign-owned capital} \\ &+ \text{Income from labor working abroad} - \text{Income earned by foreigners within home's borders}\end{aligned}$$

GNP is a more comprehensive measure of national income, but can be more challenging to measure accurately since it requires data on transactions made by citizens abroad. For most economies, the distinction between GDP and GNP is not of much consequence, except for economies which rely heavily either on remittances from abroad or on foreign capital flows for domestic economic activity[20].

[20] For instance, Ireland's GNP has been steadily declining relative to its GDP, as the country's low taxes attract foreign capital. The income generated by this capital in Ireland enters into its GDP, but because the capital is owned by foreigners, not into its GNP.

2.2.3 Gross vs Net Investment

Investment as reported by the BEA is sometimes referred to as **Gross** Investment, which includes the total spending by businesses and households on capital goods. However, since a fraction of the capital stock depreciates each period, a sizeable portion of gross investment represents the replacement of depreciated capital. **Net** investment, defined as gross investment less the depreciation of the capital stock, represents the net increase in the capital stock over a period.

$$K_{t+1} = K_t(1-\delta) + \underbrace{I_t}_{\substack{\text{Gross}\\ \text{Invt.}}} \implies K_{t+1} - K_t = \underbrace{I_t - \delta K_t}_{\substack{\text{Net}\\ \text{Invt.}}}$$

2.3 Important Macroeconomic Trends

2.3.1 US Exorbitant Privilege

A long-standing puzzle in macroeconomics is the sustained positive net foreign asset position of the United States (i.e. the fact that US GNP has exceeded GDP for a long period of time), despite its sustained negative net export position. This effectively means that the US is a net borrower from the rest of the world, and yet it receives interest income on its assets held abroad. It can be shown that this is largely driven by the higher returns earned by the US on its foreign investments than the interest paid to foreign holders of US debt, a phenomenon sometimes known as the "Exorbitant privilege" enjoyed by the US.

2.3.2 Trends in US Value Added by Sector

As the US has developed, it has undergone a process of **structural change**, wherein the industrial composition of value added has changed over time. In particular,

- The share of GDP devoted to primary activities, such as agriculture, forestry, fishing and mining has declined steadily over time.
- The share of GDP devoted to secondary activities such as manufacturing has initial risen and then declined over time. This decline has accelerated post 2000.
- The share of GDP devoted to tertiary activities, which include services such as health and education, entertainment and recreation, has steadily increased. Of particular importance is the rise in the FIRE (Finance, Insurance, Real Estate and Rental) industries.

The decline of manufacturing as a share of value added and in total employment has been a repeated cause for concern among US policymakers. The most obvious explanation for the decline is structural change - as the US becomes a richer economy, it shifts from producing manufactured goods and specializes in the production of higher value added commodities, in particular, services. Alternative explanations for this decline can be grouped into three main categories.

- **Income growth and income-inelastic demand:** This is related to the structural change point, but instead reflects preferences, not specialization. As the US has become richer, the demand for manufactured goods has not risen one for one with income due to inelastic demand. This is intuitive: it is not clear that the demand for clothing must rise one-for-one with income, for instance.
- **The impact of trade:** An increase in trade with low-wage economies has led to US firms either being out-competed by low priced imports or to these firms offshoring production, reducing the US share of manufacturing. While there is some evidence for this channel in the US, it is harder to explain the decline in some other OECD countries via the same mechanism.
- **Automation and price-inelastic demand:** Technical progress has allowed firms to produce the same level of output more cheaply, with fewer workers and more machines. Evidence that US manufacturing is becoming more productive[21] seems to provide support for this hypothesis. In order for this to work, it must be the case that demand for manufactured goods overall is relatively price inelastic - in other words, as manufactured goods get cheaper, the demand for these goods rises less than one-for-one, so the total value of manufactured goods falls relative to total output.

2.4 Growth and Business Cycles

A stylized fact about the US is that GDP per capita grows at around 2% a year, a number that has been remarkably stable for almost the entire period since 1900. However, quarter-to-quarter, GDP growth can fluctuate, giving rise to the business cycle. About 85% of quarters since the World War 2 have seen US GDP expand, and only 15% have seen declines.

A **recession** is a period of general deterioration in macroeconomic conditions. A rule of thumb for the US to be in a recession is two or more successive quarters during which real GDP has fallen, but official business cycle dates (as defined by the NBER) place weight on employment rather than on real GDP. The average postwar US recession prior to the COVID-19 pandemic recession involved a 3% decline in real GDP, a 2% decline in employment and a 2 percentage point[22] increase in the unemployment rate. Since real GDP grows at about 3% annually, a typical postwar recession represents a 6% decline in GDP relative to trend. Similarly, with employment growing around 1.5%

[21] That is, output per worker in manufacturing is actually rising.

[22] Changes in variables that are naturally defined as fractions or percentages, such as interest rates or unemployment rates, are usually reported in percentage points rather than in percentage changes. To make the distinction clear, suppose the interest rate falls from 2% to 1%. This is a 50% decrease in the interest rate, but a 1 percentage point (or a 100 basis point) decline in the interest rate.

a year typically, a recession represents a 3.5% slower growth in employment than the trend suggests.

A **Depression** is a name assigned to particularly deep and prolonged recessions. A rule of thumb sometimes used is that a depression involves a 10% decline in GDP or more. The Great Depression of 1929-1933, for instance, involved a 30% decline in GDP and a 21 percentage point rise in the unemployment rate.

The pandemic recession involved a 10% drop in US GDP from 2019Q4 to the trough (in 2020Q2). GDP fell much more deeply in European countries, particularly those dependent on tourism and related industries. While the pandemic recession was deep, the US economy and most advanced economies globally have recovered strongly in 2021, coinciding with the widespread availability of vaccines and reopening efforts, making the pandemic recession the shortest on record.

2.5 GDP and Welfare

In principle, GDP per capita is a flawed measure of welfare since it only captures the value of final goods and services available per individual in an economy, therefore ignoring a wide variety of things that individuals might consider valuable, including measures of health and safety, leisure, the quality of the environment and so on. One of the Hard Questions asks you to document that in practice, it is a useful proxy that is strongly correlated with measures of all of these factors. In particular, high GDP per capita is correlated with the extent of in-migration, suggesting that individuals do indeed believe that they would be better off in a high-income economy.

2.5.1 The Easterlin Hypothesis

Easterlin argues that individuals derive benefits primarily from *relative* consumption, i.e. that the level of utility that an individual i enjoys from consuming c_{it} at date t can be expressed as a function of the form $u(c_{it}/\bar{c}_t$ where $\bar{c}_t$ is the average level of consumption at date t across all individuals. This hypothesis rests on the empirical observation that richer individuals report being happier within countries, but average happiness rises only slowly with income across countries or over time within the same country. The hypothesis is controversial due to criticisms of the empirical work that underlies it[23].

[23]See Stevenson and Wolfers (2008) and Justin Wolfers' article here

2.5.2 The Present Discounted Value of Utility

To evaluate welfare across countries, macroeconomists sometimes use the Present Discounted Value (PDV) of utility, defined as

$$U_0 = \sum_{t=0}^{\infty} \frac{1}{(1+\rho)^t} u(c_t)$$

In this formula, U_0 denotes the discounted utility at time 0, c_t is per-capita real consumption for an individual in year t, and $u(c_t)$ is the flow value of utility from this consumption in year t. This utility at date t is discounted by the factor $1/(1+\rho)^t$, which reflects the pure time discount rate and should be interpreted as reflecting forces which make consumption today more valuable than consumption in the future.

Note that what enters the formula is not GDP per capita, but consumption per capita. The distinction between the two is crucial in some contexts - for instance, during World War II, US GDP per capita expanded enormously, but this rise was virtually entirely driven by increases in government spending, and not by consumption, which actually fell over this period.

The role of discounting is controversial, since allowing for a high discount rate effectively places less weight on future flow utility. Discount rates are typically justified by the fact that individuals might be worried about mortality - once dead, an individual clearly does not receive any further flow utility - or might just be impatient. There is also the possibility that individuals are "myopic" - they simply do not care about the future. As we will see when we study the consumption-savings decisions of households, observed levels of real interest rates are only rationalizable if individuals discount the future to some extent.

To operationalize this measure of welfare, we need to choose a flow utility function $u(c)$. A common choice is the *isoelastic* utility function,

$$u(c_t) = \frac{c_t^{1-\frac{1}{\sigma}} - 1}{1 - \frac{1}{\sigma}}$$

where $\sigma > 0$ is a parameter called the Intertemporal Elasticity of Substitution (the IES). We define the marginal utility function as

$$\frac{du(c_t)}{dc_t} = c_t^{-1/\sigma}$$

Since $\sigma > 0$, note that marginal utility declines as consumption per capita rises. Intuitively, this captures the idea that as individuals consume more, they value each extra unit of consumption less and less. However, also note that the marginal utility of consumption is always positive - that is, raising consumption can never *reduce* utility[24].

The higher σ is, the less rapidly marginal utility diminishes with consumption. When σ is close to 0, the marginal utility function approaches 0 arbitrarily quickly; when

[24] That is, the individual is never *satiated*.

σ is very large, the marginal utility function approaches a constant level of 1 for all values of c_t. One way to interpret this behavior is to think of the IES as a parameter that determines how "flexible" the consumer's marginal utility is when it comes to changing consumption per capita. If σ is high, then marginal utility doesn't change a lot when consumption changes, and the consumer doesn't care too much about varying c as long as the mean level of consumption is stable. But if σ is small, the consumer is relatively inflexible - small changes in consumption correspond to large changes in how much the consumer values a marginal unit, implying that the consumer strongly prefers a relatively constant consumption level. Thus, the higher σ is, the more willing individuals are to shift consumption over time.

The parameter σ is also inversely related to the risk aversion of the individual[25]. Under this specification of utility, note that $u''(c) = -c_t^{-1-1/\sigma}/\sigma < 0$, which implies that the utility function is *concave*, implying that individuals are always risk averse - they would always prefer to receive \$10 for sure than take a gamble that pays \$20 with probability 1/2 and nothing with probability 1/2. However, when σ is high, individuals are *less* risk averse.

Note that discounting and the impacts of the IES are distinct concepts. Discounting is a concept related purely to time, and reflects the idea that the present receives a higher weight in calculating welfare than the future does. The IES, however, is about how responsive marginal utility is to fluctuations in consumption, whether across time, across states of the world or across individuals.

2.5.3 Welfare Calculations and Comparisons

We now apply the idea of comparing the PDV of utility across individuals to studying the costs of business cycles and studying the value of economic growth. To study business cycles, we consider a stylized example. Suppose the present value of utility is

$$U = \mathbb{E}\left[\frac{c^{1-1/\sigma} - 1}{1 - 1/\sigma}\right]$$

where $\mathbb{E}$ denotes the mathematical expectations operator. Consider two worlds, A and B. In world A, suppose that in each period that c is either $1 + \varepsilon$ or $1 - \varepsilon$, each with equal probability 1/2. ε is a measure of the severity of economic fluctuations - the higher the ε, the larger the variability in consumption. U can be thought of as the value of living in this society "behind the veil of ignorance", where the individual evaluates welfare without knowing the exact trajectory of c_t that she will face. What is the present value of utility the individual faces? It is just

$$U_A = \frac{1}{2}\frac{(1+\varepsilon)^{1-1/\sigma} - 1}{1 - 1/\sigma} + \frac{1}{2}\frac{(1-\varepsilon)^{1-1/\sigma} - 1}{1 - 1/\sigma}$$

In world B, suppose that in each period c is constant and equals some level λ. Note that in this world there is no uncertainty, and the consumer knows she will get λ each

[25]The Arrow-Pratt measure of risk aversion is defined by $-cu''(c)/u'(c)$. Show that for Isoelastic utility this equals $1/\sigma$.

period with probability 1. The present value of utility the individual faces in world 2 is

$$U_B = \frac{\lambda^{1-1/\sigma} - 1}{1 - 1/\sigma}$$

We now ask: for what level of λ would the consumer be equally happy living in either world A or in world B? Since worlds A and B are identical in all other respects but for the fact that A contains business cycles and B does not, the lower λ is, the more the consumer cares about business cycles. One way to interpret λ is to note that $1 - \lambda$ is the fraction of permanent consumption that the household would be willing to give up in order to live in a world with no business cycles. To calculate λ, set $U_A = U_B$ and solve for λ to get

$$\begin{aligned} U_A &= U_B \\ \implies \frac{1}{2}\frac{(1+\varepsilon)^{1-1/\sigma} - 1}{1 - 1/\sigma} + \frac{1}{2}\frac{(1-\varepsilon)^{1-1/\sigma} - 1}{1 - 1/\sigma} &= \frac{\lambda^{1-1/\sigma} - 1}{1 - 1/\sigma} \\ \implies \lambda &= \left[\frac{1}{2}(1+\varepsilon)^{1-1/\sigma} + \frac{1}{2}(1-\varepsilon)^{1-1/\sigma}\right]^{\frac{1}{1-1/\sigma}} \end{aligned}$$

For $\sigma > 0$ it can be shown that λ decreases as ε increases - the costs of business cycles increase as they become more volatile, even though the average level of consumption doesn't change when ε rises. However, the extent to which λ falls depends on σ. When $\sigma \to \infty$ - so individuals are completely indifferent about fluctuations in consumption and care only about the mean level - it can be shown[26] that $\lambda \to 1$.

Why is this exercise important? Suppose that most individuals have a relatively small σ, as the data suggests. In this case, business cycle fluctuations in consumption are particularly costly, and this implies a role for governments to step in to mitigate how much consumption varies over the cycle through programs such as unemployment insurance even if the financing of these policies is distortionary and reduces the average level of consumption. By contrast, if most individuals have a relatively high σ, then fighting business cycles might not be as high a priority and the distortionary effects of the taxes and transfers necessary to do this might outweigh the benefits.

Now, let's consider economic growth. Suppose that

$$U_0 = \sum_{t=0}^{\infty} \frac{c_t^{1-1/\sigma}}{(1+\rho)^t}$$

Again, consider two worlds A and B. In world A, we assume that consumption grows at a constant rate g, so that $c_t = (1+g)^t$. In world B, we assume that consumption is just constant at λ. For what value of λ would the present discounted value of utility in the two worlds be the same? To interpret the answer to this question, note that λ can be thought of as the amount we would have to multiply initial consumption for the consumer to be happy to accept a world with no economic growth.

[26] Show this rigorously!

To answer this, we once again set $U_A = U_B$ and solve for λ. We have,

$$U_A = U_B$$
$$\implies \sum_{t=0}^{\infty} \frac{\left((1+g)^t\right)^{1-1/\sigma}}{(1+\rho)^t} = \sum_{t=0}^{\infty} \frac{\lambda^{1-1/\sigma}}{(1+\rho)^t}$$

We can use the fact that $(x^a)^b = \left(x^b\right)^a = x^{ab}$ to rewrite the left hand side of this expression. We get,

$$\sum_{t=0}^{\infty} \left(\frac{(1+g)^{1-1/\sigma}}{1+\rho}\right)^t = \lambda^{1-1/\sigma} \sum_{t=0}^{\infty} \left(\frac{1}{1+\rho}\right)^t$$

Noting that both sides of the equation are just infinite sums of geometric series[27], we can write

$$\frac{1}{1 - \frac{(1+g)^{1-1/\sigma}}{1+\rho}} = \frac{\lambda^{1-1/\sigma}}{1 - \frac{1}{1+\rho}}$$

where we need to assume that $\frac{(1+g)^{1-1/\sigma}}{1+\rho} < 1$. Cleaning up the expressions, we get

$$\lambda = \left(\frac{\rho}{1+\rho-(1+g)^{1-1/\sigma}}\right)^{1/(1-1/\sigma)}$$

When we stick numbers into this formula, we find that values of λ are in general large - when $\sigma = 2$, $\rho = 2\%$ and $g = 2\%$, which are plausible values empirically, we get $\lambda \approx 3.5$. This implies that to convince a consumer to abandon a world where she was guaranteed 2% growth forever, one would need to raise her constant level of consumption permanently by 350%!

Note that λ is larger when g or σ are higher and when ρ is smaller. The effect of g is intuitive - the more growth there is, the higher the present value of the stream of consumption in world A is, and the higher λ needs to be for world B to catch up. When ρ is low, the household cares about the future to a larger extent. Since the real benefits of growth come in the future, when consumption will be much larger, a low ρ raises the value of the consumption stream in world A, again requiring λ to be higher to allow world B to catch up. Finally, when σ is low, the household values growth less. This is because growth leads to consumption varying over time, which reduces marginal utility for the household in the future relative to the present. When σ is low, the household would prefer to instead have a flatter consumption profile - i.e., one with less growth in consumption but a higher level of permanent consumption - than the profile it actually gets in world A.

[27]Recall from the mathematical preliminaries that if $\beta < 1$ then $1 + \beta + \beta^2 + \cdots = 1/(1-\beta)$

2.6 Exchange Rates and Purchasing Power Parity

Most countries around the world denominate transactions in their own currency. We therefore need to use exchange rates to convert values to a common currency before comparing any nominal quantities. There are two possible choices for converting nominal values to common units: **Nominal**, or **market exchange rates** and **PPP exchange rates**.

2.6.1 Notation

Note that the exchange rate e can be reported as either

- The **foreign currency price of the domestic currency**.
- The **domestic currency price of the foreign currency**.

For instance, taking home to be the US and foreign to be the UK, the dollar-pound exchange rate can either be written as 0.71 GBP per USD if following the first convention, or as USD 1.4 per GBP if following the second. In Econ 52, unless stated otherwise, we will use the former convention: the exchange rate is always going to be the number of units of foreign currency required to purchase one unit of the home currency.

The home currency is said to **appreciate** if the exchange rate rises - that is, home's currency becomes more expensive in terms of the foreign currency. The financial press sometimes says that the home currency has “strengthened” in this case. The home currency is said to **depreciate** if the exchange rate falls - that is, home's currency becomes cheaper in foreign currency terms. Under our convention, a rise in e is an appreciation and a fall in e is a depreciation - which is hopefully more intuitive than the latter.

2.6.2 Using Market Exchange Rates

Suppose that a good sells for 200 Yen in Japan and that the market exchange rate is 100 Yen per Dollar. The price of this good converted using market exchange rates is just

$$P^{home} = P^{foreign}/e = \frac{200\ Yen}{100\ Yen/USD} = 2\ USD$$

Using exactly this idea, we could in principle convert GDP measured in foreign currency units to their equivalent dollar values, using

$$Y_{Foreign,USD} = \frac{Y_{Foreign,ForeignCurrency}}{E}$$

The idea underlying this conversion is that market exchange rates continuously adjust to ensure that the price of a good expressed in a common currency must be the same across all countries, sometimes called the **Law of One Price**. If not, in theory, there would be opportunities to earn profits by purchasing the good in countries where the exchange rates imply the good is cheap and selling them where the exchange rate implies the good is expensive. In practice, the Law of One Price does not hold - prices of goods vary widely across countries even when converted to a common currency. The different forces responsible for this include transportation costs, trade barriers including tariffs and taxes and price discrimination (where firms may charge different prices for the same good in different countries).

2.6.3 Purchasing Power Parity (PPP) Exchange Rates

Let $P^{home}, P^{foreign}$ be the average price levels in home and foreign respectively. The PPP exchange rate is defined as

$$e^{PPP} = \frac{P^{foreign}}{P^{home}}$$

For instance, if the prices of goods in Japan expressed in Yen are 150 times the prices of the same goods in the US expressed in USD, the PPP exchange rate between the Yuan and USD would be 150 Yuan per USD. PPP exchange rates reflect differences in prices across countries and are hence more accurate in reflecting differences in living standards, but suffer from the downside that they require the researcher to assume a common basket of goods across countries and that all goods are equally important in consumption baskets. Calculating PPP exchange rates requires the collection of prices of individual goods, which is expensive and time consuming. As a result, data on PPP exchange rates are usually only available with a lag, and not available for some countries[28].

2.6.4 The Real Exchange Rate

The real exchange rate is defined as the ratio of the market exchange rate to the PPP exchange rate,

$$E^{real} = \frac{E^{nominal}}{E^{PPP}} = \frac{E^{nominal}}{P^{foreign}/P^{home}}$$

A currency is said to be *at Purchasing Power Parity* with respect to another if the real exchange rate between the two currencies equals 1. In this case, the nominal exchange rate $e^{nominal}$ and the price levels at home and foreign $P^{home}, P^{foreign}$ follow the relationship

$$E^{nominal} = \frac{P^{foreign}}{P^{home}}$$

When PPP holds, goods on average have the same prices when measured in a common currency at home and in foreign.

[28]For a description of how PPP exchange rates and PPP GDP are calculated, read this Vox column and download the data here.

If $e^{real} > 1$ then the home currency is said to be **overvalued**. Note that in this case, we have $e^{nominal} > P^{foreign} / P^{home}$, so it is more expensive on average to buy goods at home prices in home currency than it is to buy them abroad at their foreign currency prices.

If $e^{real} < 1$ then the home currency is said to be **undervalued** (and the foreign currency is overvalued). Note that in this case, we have $e^{nominal} < P^{foreign} / P^{home}$, so it is cheaper on average to buy goods at home prices in home currency than it is to buy them abroad at their foreign currency prices.

2.7 Inflation and Price Levels

2.7.1 Definitions

Inflation is the change in the aggregate price level over a given interval of time. If P_t is the aggregate price level, then inflation π_t is defined by the relationship

$$1 + \pi_{t+1} = \frac{P_{t+1}}{P_t} \implies \pi_t \approx \log P_{t+1} - \log P_t$$

When prices are *falling* over time, we refer to the situation as being *deflationary*, and when inflation rates are falling over time but are still positive, we refer to the situation as being *disinflationary*. The distinction between disinflation and deflation is important to keep in mind when reading the financial press.

It is useful to distinguish between

- **Realized** Inflation, which is the actual change in prices at date t from $t-1$, as observed at end of date t.
$$\pi_t = \frac{P_t - P_{t-1}}{P_{t-1}}$$
- **Expected** Inflation, which is the expected change in prices between the present and one period from now.
$$\pi_t^e = \frac{\mathbb{E}_{t-1} P_t - P_{t-1}}{P_{t-1}}$$
where the $\mathbb{E}_{t-1}$ denotes the expectation of a random variable conditional on information available at date $t-1$.

2.7.2 Inflation Measures in the Data

The aggregate price level measure that enters the calculation of inflation is usually one of three main indicators.

- The **GDP Deflator** is a price index used to deflate GDP values to calculate real GDP. The calculation of the GDP Deflator weighs all goods by their relative shares of current expenditure, and captures the average price level across all categories of expenditure on final goods and services[29].

- The **Personal Consumption Expenditures Deflator** (PCED) is a price index that again uses current expenditure shares as weights, but only includes consumption goods, making it a good proxy for the price level faced by households regularly. The Federal Reserve targets a 2% growth in the PCE Deflator when determining monetary policy.

- The **Consumer Price Index** only measures changes in price levels of a given basket of consumption goods that is updated infrequently - on average, the weights used are about 2.5 years old. This is the measure of headline inflation often described in the financial press.

In forecasting inflation, it can help to exclude forces that cause temporary movements in inflation. Core inflation indices are constructed in the same way as regular inflation indices but exclude food and energy prices, which fluctuate due to global shocks to demand and supply. Variants of the CPI designed to achieve this include the Median CPI, which is the median rate of change in price levels across around 44 broad categories of consumer spending, and the trimmed mean CPI, which is the average change in prices for goods which excludes outliers[30]. By excluding extreme changes in prices, these measures are better suited for inflation forecasting.

2.8 Interest Rates and Asset Prices

2.8.1 Real Interest Rates

The nominal interest rate over a period is the nominal (i.e. dollar-amount) compensation earned by an individual at the end of the period if the individual lends a dollar at the start of the period. The nominal interest rate on a bond is sometimes also called its nominal yield. Since lenders typically care about the returns to their lending activity in terms of the increase in consumption they can enjoy by engaging in lending, the appropriate way to evaluate the returns to a loan is using *real* interest rates. If i_t and π_t denote the nominal interest rate between dates $t-1, t$ and the inflation rate between the same dates respectively, the real interest rate r_t between these dates is defined by

$$1 + r_t = \frac{1 + i_t}{1 + \pi_t}$$

[29] The exact method by which the GDP Deflator is calculated is involved and beyond the scope of this course. Briefly, the BEA calculates the GDP deflator as the ratio of Nominal GDP to Real GDP, where the latter is calculated using the chain-weighted quantity indices the BEA calculates using raw price and sales data for a large number of products.

[30] Typically, the trimmed mean CPI calculates the mean inflation rate for goods between the 8th and 92nd percentiles of the distribution of price changes.

This equation is sometimes called the **Fisher Equation**. Taking logs on both sides and using the approximation $\log(1+x) \approx x$ for small x we have $r_t = i_t - \pi_t$.

In practice, it is useful to distinguish between the *ex-ante* and *ex-post* real interest rates.

- The Ex-Post real interest rate $r_t = i_t - \pi_t$, where π_t is the realized inflation over the period $t-1$ to t.
- The Ex-Ante real interest rate $r_t^e = i_t - \pi_t^e = i_t - \mathbb{E}_{t-1}(\pi_t)$, where π_t^e is the expected inflation rate over the period $t-1$ to t calculated at date $t-1$.

The ex-ante real interest rate determines borrowing and lending decisions taken at any period, since the return to a loan will only be realized in the future. However, this measure cannot be calculated directly since expected inflation is not observed in the data. Instead, three ways to estimate $r^{ex-ante}$ are

- Using recently experienced core inflation, $r_t^{ex-ante} = i_t - \pi_{t-1}$. This relies on past core inflation being a good predictor of future core inflation.
- Using inflation forecasts. These can be found in the financial press or more directly from data sources like the Survey of Professional Forecasters.
- Using Treasury Inflation-Protected Securities.

2.8.2 Inflation-Protected Securities

A nominal treasury is an asset that promises the bearer the right to earn $\$1 + i_t$ at time t for every 1 that the bearer spends on buying the asset at date $t-1$. Note that the bearer's reward is fixed in nominal terms - that is, she earns $\$1 + i_t$ irrespective of the inflation rate. Clearly, the ex-ante real return to this bond r_t (approximately) satisfies the Fisher equation $r_t = i_t - \mathbb{E}_{t-1}\pi_t$.

An Inflation-Indexed Security is an asset that works as follows:

- Over the period $t-1$ to t, you receive a fixed nominal rate of return r_t^{TIPS}.
- At the end of the period, the bond's principal resets to $1 + \pi_t$, where π_t is realized inflation over the period.
- The bond's total nominal return is therefore $1 + r_t^{TIPS} + \pi_t$.

Clearly, the ex-ante real return to this asset is $1 + r_t^{TIPS} + \mathbb{E}_{t-1}\pi_t - \mathbb{E}_{t-1}\pi_t = 1 + r_t^{TIPS}$.

Suppose investors are risk neutral - that is, they only care about the expected returns on an asset, and not about how risky that asset's returns are. If so, the returns on the two assets must be equal, and so

$$r^{ex-ante} = i_t - \mathbb{E}_{t-1}\pi_t = r_t^{TIPS}$$

which allows us to estimate the ex-ante real interest rate. In practice, since inflation-indexed securities are fully insured against (official) inflation risk - note that the real return is just r_t^{TIPS}, which is fixed - they earn a risk premium, which means that $r_t^{TIPS} < r^{ex-ante}$. The inflation rate $\pi^{BR} = i_t - r_t^{TIPS}$ is sometimes called "Breakeven inflation" and provides an estimate of inflation expectations.

Modern monetary policy involves setting nominal interest rates in order to influence ex-ante real interest rates, which are the variables that actually affect the spending and saving choices made by households.

2.8.3 Asset Pricing

Interest rates can be used to convert cash received at different points in time to a common unit. The reason this is important is because whenever interest rates are nonzero, money has "time value" - \$1 received today is worth \$$1 + i_t$ tomorrow, which is greater than the value of \$1 tomorrow. In particular, \$1 received at time $t+1$ is only worth \$ $\frac{1}{1+i_{t+1}}$ today, which means that the price of a promise to receive \$1 tomorrow should be worth only \$ $\frac{1}{1+i_{t+1}}$ today.

We can extend this logic further. Suppose the nominal interest rate was constant, at i per year. How much would you pay **today** in order to receive \$1 **in** t **years**? Well, a dollar today, with interest compounded annually, is worth $(1+i)^t$ dollars t years from now. A dollar received t periods from now is worth, well, \$1 t years from now. This is like saying a dollar t years from now is worth only $1/(1+i)^t$ dollars today. Therefore, you should never pay more than

$$Q_{0,t} = \frac{1}{(1+i)^t}$$

The ratio $\frac{1}{(1+i)^t}$ is the **Present Value** of \$1 in t-periods, representing how much a dollar in t years is worth *in units of dollars today.* Notice that $Q_{0,t}$ acts a lot like a nominal exchange rate converting dollars received at date t to their value in date-0 dollars, and is sometimes called the intertemporal price of a claim to \$1 at date t. We can use this insight to price any asset we want, as long as we know the value of the payments we receive from the asset at any date.

How does this work? Suppose you have a stream of dollar payments starting tomorrow worth $X_1, X_2, X_3, \dots$. How much is this stream of payments worth today (date 0)? We first need to convert the stream into comparable units - i.e. present values - and then we can just add them up.

$$Q_0^X = Q_{0,1}X_1 + Q_{0,2}X_2 + Q_{0,3}X_3 + \dots = \frac{X_1}{(1+i)} + \frac{X_2}{(1+i)^2} + \frac{X_3}{(1+i)^3} + \dots$$

Each of the quantities $Q_{01}X_{01}, Q_{02}X_{02}, Q_{03}X_{03}$ are now expressed in the same units - dollars in period 0. Thus, Q_0^X is the total value in period-0 \$ of the claim to the payment stream $\{X_t\}$. It would not make sense for a buyer of this asset to pay more than Q_0^X for it, and it would not make sense for a seller to charge a price lower than this (since the seller could always receive the amount Q_0^X in present value by just not selling the asset). Thus, the price of the asset must be Q_0^X. This formula can be used to calculate the price of *any* asset you want, as long as you know what the X's are.

Let's go through a couple of examples to see this in action. First consider the simplest asset, the *Consol*, which is a bond that works like this:

- At date 0 (today) you pay the bond's price Q_0^{Consol} to buy it.
- Starting at date 1 and then forever, you get \$1 each period, so $X_1^C = X_2^C = \cdots = 1$

Applying the Present Value formula, we have

$$\begin{aligned} Q_0^{Consol} &= Q_{0,1} \times 1 + Q_{0,2} \times 1 + Q_{0,3} \times 1 + \dots \\ &= \frac{1}{(1+i)} + \frac{1}{(1+i)^2} + \frac{1}{(1+i)^3} + \dots \\ &= \frac{1}{(1+i)} \left[1 + \frac{1}{(1+i)} + \frac{1}{(1+i)^2} + \dots \right] \\ &= \frac{1}{(1+i)} \times \frac{1}{1 - \frac{1}{1+i}} = \frac{1}{i} \end{aligned}$$

where the final line uses the formula for the sum of an infinite geometric series with common ratio less than 1.

How does the price of a consol vary with Inflation? Recall that from the Fisher equation, $i \approx r + \pi$, so

$$Q_0^{Consol} = \frac{1}{i} \approx \frac{1}{r + \pi}$$

Note that given a real interest rate, higher inflation *reduces* bond prices. This is because the amount paid by the bond is fixed in *nominal* terms, and hence as inflation reduces the purchasing power of a dollar - you can buy less goods with each dollar - the real return to saving in the consol falls.

Next, consider a hypothetical asset we'll call a "TIPS Consol", which works like this:

- At date 0 (today) you pay the bond's price $Q_0^{TIPS,Consol}$ to buy it.
- Starting at date 1 and then forever, you get a payment of \$ $X_t^{T,C}$ where $X_t^{T,C} = (1+\pi)^t$.

The present value formula gives

$$\begin{aligned}
Q_0^{TIPS,Consol} &= \frac{1+\pi}{1+i} + \frac{(1+\pi)^2}{(1+i)^2} + \frac{(1+\pi)^3}{(1+i)^3} + \dots \\
&= \frac{1}{1+r} + \frac{1}{(1+r)^2} + \frac{1}{(1+r)^3} + \dots \\
&= \frac{1}{1+r}\left[1 + \frac{1}{1+r} + \frac{1}{(1+r)^2} + \dots\right] \\
&= \frac{1}{1+r}\left[\frac{1}{1-\frac{1}{1+r}}\right] \\
&= \frac{1}{r}
\end{aligned}$$

Note that the real return to the TIPS is r, and this is independent of inflation. This happens because the nominal amount paid on the TIPS rises with inflation one for one.

3 Production and The Labor Market

We first describe a simple macroeconomic framework to think about production. We then study the first of the main markets in macroeconomics: the market for labor. We will study the neoclassical approach to the labor market, and conclude with an introduction to a search model.

3.1 Modeling Production

3.1.1 The Production Function

A **Production function** links quantities of factor inputs - labor and capital - to the quantity of final output produced in an economy. Mathematically, the production function can be written as

$$Y_t = A_t F(K_t, N_t)$$

where

- Y_t is a measure of final output, usually Real GDP
- K_t represents the stock of physical capital used in production
- N_t is a measure of labor input, which is total hours worked in the economy
- A_t is a measure of the efficiency with which output is produced from a given amount of factors of production in the economy, and is called **Total Factor Productivity (TFP)**.

Different production functions imply different relationships between factor inputs and final output. Two workhorse production functions we will see are

- The **Cobb-Douglas** Production Function:

  $$Y_t = A_t K_t^{\alpha} N_t^{1-\alpha} \qquad 0 < \alpha < 1$$

  As we'll see, this production function has some desirable properties with respect to the data. We'll show shortly that the share of final output that goes to labor when a firm using this production function maximizes profits is a constant, which is, remarkably, true in the data.

- The **Constant Elasticity of Substitution** Production Function:

  $$Y_t = A_t \left[\omega K_t^{\rho} + (1-\omega) N_t^{\rho}\right]^{1/\rho} \qquad \rho \geq 0, \omega \in [0,1]$$

  The CES production function is a generalization of many other popular production functions, including the Cobb-Douglas, making it a popular choice for quantitative exercises and empirical work.

3.1.2 Properties of Production Functions

Some properties of the production function are the following.

- **Returns to Factors**: Define the **Marginal Product** of a factor as the change in output for a given change in that factor's input, holding inputs of all other factors constant. Thus, the marginal products of labor and capital, respectively MPN and MPK, are given by

$$MPN = \frac{\partial AF(K,N)}{\partial N} \qquad ; \qquad MPK = \frac{\partial AF(K,N)}{\partial K}$$

  We will typically assume that MPK declines as K rises, and that MPN declines as N rises.

- **Returns to Scale**: Say we multiply inputs of labor and capital by a factor $\lambda > 1$. A production function is said to have
  - **Increasing** returns to scale, if output rises by more than a factor λ,
  - **Constant** returns to scale, if output rises by exactly a factor of λ,
  - **Decreasing** returns to scale, if output rises by less than a factor λ

- **Complementarity between K, N:** Capital and Labor are said to be **complements** if a higher level of capital raises the marginal product of labor and a higher level of labor raises the marginal product of capital.

Note that the Cobb-Douglas production function has diminishing returns to each factor input individually, constant returns to scale in capital and labor combined, and features complementarity between K, N. In addition,

- A higher value of A raises both MPK and MPN.
- The elasticity of output to each factor equals that factor's exponent. That is,

$$\varepsilon_K^Y = \frac{d \log Y}{d \log K} = \alpha \qquad \varepsilon_N^Y = \frac{d \log Y}{d \log N} = 1 - \alpha$$

3.1.3 Factor Shares and Elasticities for Cobb-Douglas

An important property of the Cobb-Douglas production function is that under perfect competition, it implies that the shares of income that go to capital and labor are constant irrespective of wages or interest rates. We now prove this result by solving the firm's problem. Note that we will repeat this exercise more generally below, so re-reading this subsection after going through that section is a good idea.

Consider a firm that maximizes profits, by solving the maximization problem

$$\max_{K,N} PAK^{\alpha}N^{1-\alpha} - WN - RK$$

where P, W, R are the price of final goods sold by the firm, the nominal wage rate and the nominal rental rate of capital respectively[31]. The firm's first order conditions with respect to K, N are

$$P\alpha AK^{\alpha-1}N^{1-\alpha} = R$$
$$P(1-\alpha)AK^{\alpha}N^{-\alpha} = W$$

Multiply the first equation by K and the second equation by N to get

$$P\alpha AK^{\alpha}N^{1-\alpha} = RK$$
$$P(1-\alpha)AK^{\alpha}N^{1-\alpha} = WN$$

Observe that the term $AK^{\alpha}N^{1-\alpha} = Y$ by definition of the production function. Thus, we get

$$\alpha PY = RK \implies \frac{RK}{PY} = \alpha$$
$$(1-\alpha)PY = WN \implies \frac{WN}{PY} = 1-\alpha$$

which shows that the factor shares of total income are independent of the amount of output made or of the inputs of the two factors.

Finally, there is one more useful property of the Cobb-Douglas production function. Adding the two equations, we get

$$pY = RK + WN$$

which states that the total revenue of the firm must equal the total payments made to labor and capital[32].

In the data, the labor share of income has been declining since 2000, from around 65% to about 60%. This decline is concentrated in manufacturing and has been seen in other countries but not in some others (for instance, it is visible in China but not in Europe). The most common explanation for the declining labor share are

- changes in technology involving automation (which one can think of as the impact of rising α in the Cobb-Douglas production function)
- rising market power (which raises the share of income that goes to pure profits)
- globalization, including the outsourcing of labor-intensive tasks, which raises the capital intensity of domestic production.

[31] As we'll see later, a more general model of the firm will replace R by the nominal *user cost of capital*.

[32] This is an example of Euler's theorem for functions that are homogeneous of degree 1. Applied to production, it states that if a general production function $F(X_1, X_2, \ldots, X_M)$ involving the M factors of production $X_1, \ldots, X_M$ satisfies constant returns to scale, then

$$F(X_1, \ldots, X_M) = \sum_{i=1}^{M} X_i F'(X_i)$$

3.1.4 TFP and Labor Productivity

Recall that the typical production function can be written as

$$Y = AF(K, N)$$

Economists use two main measures of productivity:

- Labor Productivity, defined as output per worker, $Y/N = AF(K, N)/N$. If production is Cobb-Douglas, then

$$\frac{Y}{N} = A\left(\frac{K}{N}\right)^{\alpha}$$

- Total Factor Productivity, defined as $Y/F(K, N) \equiv A$.

Note that both TFP and labor productivity depend on an economy's technology (including the quality of capital, the quality of infrastructure (if infrastructure is not included in the capital stock already), and the institutional arrangements surrounding production, including the nature of competition, regulation, management and the allocation of resources across firms). The key distinction between the two measures is that labor productivity depends on the capital to labor ratio (a higher K/N implies a higher Y/N, all else equal), while TFP does not.

What determines productivity?

- **Technology:** It can be shown that technical progress, the ability to make more output than before using the same amount of labor and capital, can explain about 20% of the level of A. This manifests itself in the rapidly declining prices of new technologies, for instance.
- **Human Capital:** This includes factors like the quality of skills, experience and the level of education possessed by workers. Experience and education account for about 20% of the variation in worker wages, and each extra year of education and experience raise wages about 10% and 3% respectively, suggesting substantial productivity gains from raising either margin of human capital. Schooling attainment, the average years of education possessed by workers, has been trending up in most countries, and can explain about 1/3 of observed growth in A.
- **Infrastructure:** Measures of capital stock available in the data can sometimes exclude the value of public capital, which includes the value of roads, utilities, port facilities, and other forms of infrastructure. In this case, the productive benefits of this public capital can show up in the productivity term A.
- **Competition:** When firms compete either with other domestic firms or with foreign players (when trade is freer or entry barriers to an industry are reduced), measured productivity tends to rise as firms are under more pressure to exploit efficiencies and reduce costs. However, there is some evidence that higher competition can inhibit innovation by reducing the rewards to a successful new product or process, which can affect the rate of TFP growth in turn.

- **Management:** A relatively new strand of literature, pioneered by Stanford's own Nick Bloom and co-authors, examines the role played by managers and management practices in determining productivity at the firm level, and shows that globally the average quality of management practices co-moves with measured TFP.
- **Misallocation:** Any forces in an economy which prevent the appropriate assignment of capital and labor across firms in an economy will reduce measured TFP, and we will discuss this in more detail when we study facts surrounding economic growth.

3.1.5 Productivity Trends

Since the 1950s, US labor productivity growth has been through several phases, starting with a growth slowdown in the postwar period until around 1973, when growth accelerated. The period after 2004 has seen a prolonged stagnation in productivity growth, leading prominent economists including Larry Summers and Robert Gordon to worry about the prospect of "secular stagnation."

There are a number of considerations when studying long run trends in productivity growth that can temper this conclusion.

- Since the 1990s a rising share of the US economy is in the technology sector, which has seen large increases in productivity. However, it is in general more difficult to measure the value of the goods and services produced by this sector[33]. Thus, it is possible that the productivity growth slowdown is an artifact of mismeasurement.
- It is possible that inflation is overstated by the way current price indices are computed. For instance, it is very difficult to account for improvements in the quality of goods and even more so for services, which are a rising share of the economy. If goods are improving in quality, it is as if a consumer purchasing a unit of the good is receiving "more value per good" - in which case, changes in the price of the good itself overestimate changes in the cost of delivering a unit of value to the consumer[34]. This bias may be leading to an underestimation of the growth rate of real GDP and hence of labor productivity[35].
- Price indices have a hard time dealing with the introduction of new varieties, since it is difficult to assign them a base-year price, and also have a hard time dealing

[33]How would you value having access to Gmail? Since Gmail is free, no GDP-entering transactions occur when we use this service.

[34]To make this concrete, consider computer prices. Suppose that between dates t and $t+1$ the average computer price is \$100 in both periods, but the average computer grows from being able to do 1000 operations per second to 10000 operations per second. Then the price per unit of performance performed by the computer - measured by operations per second - has actually *fallen* ten-fold, so if we were to measure computing costs using the flat \$100 per machine, we would be overstating inflation in terms of what consumers actually care about ten-fold as well!

[35]The discussion between Russ Roberts and Susan Houseman on the EconTalk podcast here covers many of these issues.

with goods which are "free"[36]. This would lead to the problem of overestimation of inflation being even worse.

A counterargument to these issues is that for them to explain the slowdown in productivity, these problems would have had to have gotten worse over time - that is, it would have to be the case that the overestimation of inflation has gotten worse since the mid 2000s. There is little evidence that this is the case.

3.2 Labor Demand

The neoclassical model of labor demand posits that the choice of how much labor to hire at each wage rate is the result of firms choosing their labor input to maximize profits. To see how this works, consider a firm that operates the technology

$$Y = AF(K, N)$$

and chooses labor and capital N, K to maximize profits,

$$\pi(K, N; w, r) = PAF(K, N) - WN - rK$$

We will assume that the technology $AF(K, N)$ satisfies diminishing marginal products of capital and labor. Note that P is the price of final output produced by the firm, W is the nominal wage rate and r is the rental cost of capital. This is an unconstrained maximization problem, and the first order condition with respect to labor[37] gives

$$P\frac{\partial AF(K, N)}{\partial N} = W \implies MPN = \frac{W}{P}$$

MPN denotes the Marginal Product of Labor, which (recall) is the increase in output the firm could produce by hiring an extra hour of labor input. To interpret the condition $MPN = W/P$, first recall that we assumed that MPN declines as N increases, given capital stock[38] K. The condition states that the firm should continue to hire workers until the cost to the firm of hiring the marginal hour worked - the real wage, W/P - just offsets the benefit to the firm, which is the extra output that would be produced by the final worker - which is the MPN.

The plot of the marginal product of labor against the amount of labor represents the firm's Labor Demand curve. Note that

- The labor demand curve is downward sloping, since the lower the real wage, the lower the marginal product of labor has to be to justify hiring the marginal worker, and the higher the level of labor associated with the lower MPN since MPN decreases in N.

[36]How would you value having access to Gmail? Since Gmail is free, no GDP-entering transactions occur when we use this service.

[37]We will discuss the FOC with respect to Capital in great detail in the next section.

[38]Intuitively, the more workers are hired, the more crowding of workers on a given unit of capital occurs and the less capital per worker there is, making each worker on the margin less productive.

- Increases in A and in K lead to upward/rightward shifts of the labor demand curve, since they raise MPN for each level of labor input and make hiring each unit of labor more profitable. Note that when A and K go up, firms respond by raising the amount of labor they hire at each wage rate, implying that they produce more output as well - this is consistent with profit maximization, where firms respond to more productive, and hence more profitable, labor by selling more output[39].
- Note that our model is that of a competitive firm that takes prices W, P as given. To determine the equilibrium consequences of changes in A and K, we cannot use the labor demand curve alone - we must combine it with a model of labor *supply*, which we now turn to.

3.3 The Static Consumption-Leisure Trade-off

Consider a consumer deciding between the number of hours to work and the amount of consumption she wants. For each hour the consumer works, she earns \$ W, which she can use to purchase W/P units of consumption. The consumer solves the problem

$$\max_{C,N} U(C,N) \qquad \text{subject to} \quad (1+\tau_C)PC = (1-\tau_N)WN$$

where

- τ_N is the marginal tax rate on labor income, and captures ordinary income taxes, payroll taxes and transfers that are conditioned on income like the earned income tax credit.
- τ_C includes all taxes on consumption, which can include sales taxes or value added taxes.

We will generally assume that

- $\frac{\partial U}{\partial C} > 0, \frac{\partial U}{\partial N} < 0$: the consumer always prefers more consumption to less, and always dislikes more hours worked.
- $\frac{\partial U}{\partial C}$ is decreasing in C: that is, there is *diminishing marginal utility from consumption*. The consumer's marginal benefit from an extra unit of consumption is larger when her current consumption is lower. Intuitively, the utility gained from an extra bite of food is much larger when you just start eating than when you're already pretty full.

[39] Where does the demand for this extra output come from? It is important to note that this model of the labor market is a partial equilibrium model, and that all forces that determine the demand for goods are captured in the price level P that the firm takes *as given*. When we "close" the model, we must add in a model of the demand for goods that determines the price level.

- $\frac{\partial U}{\partial N}$ is decreasing in N, i.e. becoming larger in absolute value, but more negative. The pain of an extra hour worked is much smaller at 9:00 AM when you're just starting to work than it is at 6:00 PM, when you really want to go home after a long day.

The Lagrangian for this problem is

$$\mathcal{L}(C,N) = U(C,N) + \lambda(WN - PC)$$

The First-Order Conditions are

$$\frac{\partial \mathcal{L}}{\partial C} = U_C(C,N) - (1+\tau_C)\lambda P = 0 \implies U_C(C,N) = \lambda(1+\tau_C)P$$
$$\frac{\partial \mathcal{L}}{\partial N} = U_N(C,N) + \lambda(1-\tau_N)W = 0 \implies -U_N(C,N) = \lambda(1-\tau_N)W$$

Combining the FOCs to eliminate the Lagrange Multiplier λ,

$$-U_N(C,N) = \frac{(1-\tau_N)W}{(1+\tau_C)P} U_C(C,N)$$

The left side of this static FOC is the marginal cost of working an extra hour for the household, in utils. The right side of the FOC captures the marginal benefit of an extra hour worked. To see this, note that an extra hour worked raises consumption by $\frac{(1-\tau_N)W}{(1+\tau_C)P}$ units, each of which raises utility by $U_C(C,N)$ utils, for a total gain of $\frac{(1-\tau_N)W}{(1+\tau_C)P}U_C(C,N)$. The FOC thus captures the intuitive idea that the household should work until the marginal pain of an extra hour worked is just offset by the marginal gain in terms of extra consumption the household can now afford.

What determines labor supply?

- The after-tax real wage: any changes in taxes or in the real wage itself will affect the relative benefit of working an hour and affect labor supply by each individual.
- The working-age population: Our model studied the response of an individual to changes in their individual consumption and wage rates. The response of total labor supply is the product of individual labor supply and the population of individuals who can work.
- The marginal utility of consumption, which is pinned down by the **Present Value of Lifetime Resources**.

3.4 The Present Value of Lifetime Resources

Define the Present Value of Lifetime Resources ($PVLR$) by

$$PVLR \equiv \sum_{t=0}^{\infty} \left(\frac{1}{1+r}\right)^t [Y_t - T_t + Tr_t]$$

where

- Y_t is total pre-tax-pre-transfer income earned by the household from all sources. In particular, it includes both income from labor and capital at date t.
- T_t is total taxes paid by the household at date t.
- Tr_t is total transfers received by the household at date t.

As we will show in the section on savings REFERENCE, the consumption chosen by a consumer who has access to perfect credit markets where she can save or borrow as much as she likes at the common interest r will depend *only* on the $PVLR$, and not on current income (apart from the effects changes in current income have on $PVLR$). In particular, if we change current income but change future income in such a way that $PVLR$ remains unchanged, current consumption will not change at all.

Since current consumption depends only on $PVLR$, so does the marginal utility of consumption, $\frac{\partial U}{\partial C}$. Thus, in the labor-leisure trade-off, it is changes in $PVLR$ that affect the term $U_C(C, N)$ on the right side, not just changes in current income.

3.5 Income and Substitution Effects

To understand the economics of the labor-leisure trade-off in more detail, it is useful to decompose the response of labor supply to changes in the labor market into two components.

- The **Substitution** effect is the change in labor supply that corresponds to a change in the wage rate, holding the household's consumption fixed[40].
- The **Income** effect is the change in labor supply that corresponds to a change in the household's consumption (and by the previous section, in the household's $PVLR$), holding the wage rate fixed.

To gain some more understanding of the nature of the two effects, consider the case of an individual who currently earns about \$22 an hour[41]. Suppose this individual is suddenly offered a new job that pays \$1000 an hour. What should go through the individual's mind?

- For every extra hour the individual works, its consumption rises by *\$1000*. That is a lot of money - probably enough to induce the individual to work more hours to raise consumption, given the current number of hours worked and current consumption level. This is the **substitution** effect at work - effectively, each hour of leisure now costs the consumer \$1000 in foregone consumption, making leisure expensive enough that the individual wants to substitute toward labor supply.

[40]More technically, the substitution effect requires holding the marginal utility of consumption for the household fixed, which as you saw above is equivalent to holding $PVLR$ fixed. This is the change in the Hicksian labor supply (or leisure demand).

[41]This roughly corresponds to the average nominal hourly wage of production or non-supervisory employees in the US currently. See FRED series AHETPI.

- For every extra hour the individual works, its consumption rises by *\$1000*. That is a lot of money - in fact, if this individual were to work just 1 hour, she'd make more money than in an entire 40 hour workweek at her old job. Thus, the individual is in effect a lot richer, and this should induce her to want to consume more leisure[42]. This is the **income** effect: the desire to consume more leisure at higher income levels will tend to reduce labor supply.

Note that

- A given change in the labor market may have no effects, a pure income effect, a pure substitution effect or represent a combination of both effects, and figuring out which effects occur is an important part of solving problems involving the labor market.
- An income effect typically involves a shift in the labor supply curve, while a substitution effect typically involves a movement along the labor supply curve.

The evidence from looking at labor supply across countries and within countries over time suggests that in the long run, income effects typically dominate substitution effects.

3.6 Shocks in the Labor Market

When we study changes in the labor market, it is important to keep track of the following.

- Is the change **permanent**, i.e. does the change affect the long-run level of income, and therefore affect $PVLR$? Or is it **transitory**, and therefore unlikely to affect $PVLR$ directly?
- Are we studying short run changes - in which capital cannot change - or long run changes, in which capital can adjust? If we are studying the long run, what happens to the marginal product of capital in the long run? We will study the determinants of the capital stock in the long run in more detail in the next section, but for now, it suffices to note that in the long run, profit maximization by firms requires that the marginal product of capital MPK equal the user cost of capital, a required rate of return that is increasing in the interest rate, the rate of capital depreciation and the tax rate on profits. In the long run, any shock that leaves these three factors unchanged will ensure that the user cost of capital is constant, which in turn will force MPK to remain unchanged. Note that the MPK is pinned down by TFP A and the K/N ratio. Thus, in the long run, if TFP is constant, this will imply that the long-run K/N ratio is constant as well.

[42]Technically, this assumes that leisure is a "normal good" - that increases in income induce an increase in leisure demand, a fact which is borne out by global evidence that societies with higher hourly wages also tend to have lower hours worked per worker.

Let's consider an example: an increase in immigration that raises labor supply permanently. Assume that production is Cobb-Douglas, immigration doesn't affect the user cost of capital, TFP or the stock of installed capital. In the short run (with a fixed capital stock),

- Immigration won't affect the labor demand curve at all, since it doesn't change any of the determinants of MPN.
- The labor supply curve shifts to the right due to the direct effect of the new immigrants, reducing wages.
- However, on top of this direct effect, since the new immigrants raise labor supply permanently, the wage reduction is also permanent - so we expect $PVLR$ to decline. The income effect now implies a further increase in hours worked, and so in equilibrium, there is a further decline in the real wage and a further increase in employment in the short run.
- In the long run, since there's no change in TFP and no change in the user cost of capital, there is no change in the K/N ratio. However, we argued that there is an increase in labor supply, driven by the change in $PVLR$ and the increased number of workers. Thus, it must be the case that the capital stock rises to the point where the K/N ratio has risen back to its original level. This increase in the capital stock will shift the labor demand curve out to the right, leading to an increase in the real wage in the long run.
- Where will this process stop? We know that the new K/N ratio goes back to the original level, and that TFP is unchanged. Since the MPN under Cobb-Douglas is just $A(1-\alpha)(K/N)^{\alpha}$ we know that the MPN is back to its original level as well. But for the firm to be maximizing profits, we know that MPN equals the real wage - so the real wage must return to its original level as well!

3.7 A Search and Matching Approach

The neoclassical labor market studies the number of hours worked and the number of hours employed by firms, but doesn't have anything to say about involuntary unemployment, which is a situation where workers who are able and willing to work at the current wage rate firms are paying are unable to find jobs. There are several factors that the model ignores that can explain the existence of unemployment; unfortunately, it can be difficult to tell these apart in the data and in models.

- **Structural** unemployment, which is a form of permanent, long-term unemployment, occurs due to issues like mismatches between skills and job requirements and technological changes that make certain skills obsolete.
- **Frictional** unemployment occurs due to the fact that it can take time for a worker to find a firm that is a good match for them. While this form of unemployment is typically temporary, it is a key reason for average unemployment rates to be positive.

- **Cyclical** unemployment is associated with fluctuations in the demand for goods, which leads firms to either raise or lower the demand for labor over the cycle. Cyclical unemployment can only exist if there are frictions that prevent real wages from fully adjusting to equate labor demand and supply.

Data on US employment typically comes from two sources: the Establishment Survey of around 400,000 businesses which covers non-farm employment of production workers, and the Household Survey which covers about 60,000 households and tracks the share of respondents who are unemployed. The establishment survey is the source for headline jobs numbers, but misses individuals in the farm sector and the self-employed. The household survey includes all workers, both production and non-production, and is useful for tracking the unemployment rate directly. In recent years, more granular data on job vacancies and turnover in labor markets is available from the Job Openings and Labor Turnover Survey (JOLTS).

We now sketch a simple accounting framework to think about unemployment. Let U and E denote the total number of unemployed and employed individuals in an economy. We define the labor force as the sum $E+U$. Let pop denote the working-age population in the economy. Define

- **The Unemployment Rate** $u = \frac{U}{U+E}$. Note that the denominator is the labor force, not the population. The employment rate is just $e = 1 - u$.
- **The Labor Force Participation Rate** $LFPR = \frac{E+U}{pop}$.
- **The Employment-Population Ratio** E/pop.

Note that the Employment-population ratio is not equal to the employment rate since the denominator of the employment rate includes only workers who are in the labor force. The difference between the labor force and the working age population is the set of workers sometimes called "discouraged" workers, who are unwilling to even search for a job at present.

For simplicity, ignore the participation margin and suppose that there are only two groups of workers, the employed and the unemployed. Let d_t and f_t denote the rates at which jobs are destroyed - that is, the share of employed workers who leave their jobs due to fires or quits in a period - and the job finding rate, the share of unemployed workers who find a job each period. Suppose at date t there are E_t and U_t employed and unemployed workers respectively. Then between dates $t, t+1$,

- $d_t E_t$ workers leave the employment pool and enter unemployment.
- $f_t U_t$ workers leave the unemployed pool and become employed.

Thus, the number of employed and unemployed people at $t+1$ must satisfy

$$E_{t+1} = (1 - d_t)E_t + f_t U_t U_{t+1} \qquad = (1 - f_t)U_t + d_t E_t$$

Suppose d_t, f_t are constant over time and equal to d, f respectively. Then,

- The unemployment rate eventually settles down to $u = \frac{d}{d+f}$
- The average duration of employment equals $1/d$ and the average duration of unemployment equals $1/f$.

Modern theories of involuntary unemployment study the determinants of f and d and how these rates vary over the cycle. Typically, we see that the job finding rate

- increases in the ratio of vacancies to the number of unemployed workers, since this implies that each unemployed worker is more likely to match to a vacancy
- increases in the relative size of wages to unemployment benefits, since higher wages induce more search effort by unemployed workers. Note that the ratio of benefits to previous wages is sometimes called the **replacement rate**.
- increases in the extent of goods demand, which can affect labor demand by firms. This force is typically cited as an explanation for cyclical unemployment.

The job destruction rate is typically seen to decrease with the extent of firing costs and the level of goods demand, and increase as real wages relative to worker productivity rise (this can happen when prices or wages are sticky) and with technological changes or changes that induce firm destruction. An increase in firing costs may also impact job finding rates by reducing the incentives of firms to create vacancies - if firms fear high costs of laying off workers, they may be more reluctant to hire them in the first place.

In recessions, increases in unemployment are largely driven by declines in job finding rates f and not by higher job destruction rates. Thus, the appropriate way to interpret unemployment in recessions is not in terms of a long period during which layoffs are high, but rather as a period during which firms stop hiring workers. Thus, even given a constant job destruction rate, the unemployed "pile up" in the unemployment pool since the outlet into employment, the job finding rate, becomes much smaller.

4 Consumption, Savings and Investment

Consumption and Private Sector Investment make up about 67% and 17% of GDP, making understanding the determinants of each crucial in any macroeconomic framework. In this section, we review some theories of consumption and savings behavior, understand the neoclassical theory of investment, and describe equilibrium in the capital market.

4.1 Setting the Stage

Recall from Section 1 that total output, using the Expenditure approach, can be written as

$$Y = C + I + G + NX$$

For simplicity, let's assume that the economy is closed, so that $NX = 0$. Define *national savings* S as

$$S = Y - C - G$$

That is, national savings equals national income minus total current spending by households and governments. Combining these two equations, we get

$$S = I$$

which is an important identity.

In the models we will study in this class, there will be only one asset - the capital stock that firms operate and households own. As a result, households will be able to save only by accumulating capital. This may seem strange if you are used to the idea of saving in, say, a checking account at a bank, or by buying government bonds. But the two ideas are in fact completely consistent with each other.

To see why this is, note that most assets that households purchase are in *zero net supply* - the creation of the asset creates an offsetting liability in the economy somewhere. When a household saves by depositing some money at a bank, it acquires an asset - the value of the deposit. But at the same time, the bank at which the deposit is stored acquires a liability - since households are allowed to withdraw deposits any time they want, the bank - more precisely, the bank's owners - effectively owe the value of the deposit to the household. The net effect of this transaction on total saving by all agents in the economy (the depositing household and the bank owners) is zero.

Consider the purchase of a government bond. The bondholder acquires an asset, the bond, but there is an offsetting liability created in the economy since the government's debt goes up by exactly the same amount. Since all government debt is effectively owed by taxpayers, this transaction does not change the net saving of the entire country either.

When will a transaction actually lead to the creation of net positive saving? Only when it leads to the creation of an asset which is not in zero net supply, which means that the creation of such an asset must not generate an equivalent offsetting liability somewhere else. But for any asset to generate positive value, it must provide a net increase in productive services at present or in the future - and the set of assets that satisfy this condition is the capital stock.

Another way to think of this is to trace the path of a dollar saved. A dollar saved by the household at a bank will be used by the bank to, say, make a mortgage loan. The mortage loan is an asset for the bank, but the offsetting liability belongs to the borrower, so the mortgage loan is also an asset in zero net supply. The borrower uses this money to pay a home construction company to finance the creation of a new house. At this point, a new asset has been created - a house, which will provide residential services, which will be included in national income (either in the form of imputed rents on the house if the borrower chooses to live in the house or in the form of rental income if it is rented out). The overall transaction, from the point of the aggregate economy, is the household choosing to spend a dollar of saving on accumulating housing capital - the intervening change of assets all cancel out.

Note that the fact that residential mortgages are assets in zero net supply does not mean that the intervening chain of assets is unimportant from the point of view of analysing the behavior of the macroeconomy. Since the financial crisis of 2007-08, understanding frictions in the intermediation of savings to credit has taken center stage in macroeconomic modeling.

4.2 The Consumption-Savings Trade-off

4.2.1 The Keynesian Consumption Function

One of the earliest attempts to understand consumption was proposed by John Maynard Keynes, who argued that consumption should be thought of as a linear function of current income. Letting C_t and Y_t denote current consumption and current income respectively, Keynes' idea can be formalized in the *Consumption Function*

$$C_t = a + bY_t$$

Define the Marginal Propensity to Consume (the MPC) as the change in consumption for a marginal change in income,

$$MPC_t = \frac{dC_t}{dY_t}$$

The MPC can be interpreted as the fraction of a marginal increase in income that is consumed. It is convenient to define the Marginal Propensity to Save (the MPS) as the change in saving for a marginal change in income,

$$MPS_t = \frac{dS_t}{dY_t} = \frac{d(Y_t - C_t)}{dY_t} = 1 - MPC_t$$

Note that under the Keynesian Consumption function, ww have $MPC_t = dC_t/dY_t = d(a + bY_t)/dY_t = b$, which is a constant independent of current income or wealth.

Empirically, the Keynesian consumption function struggles with the fact that there is wide variation in estimated marginal propensities to consume.

4.2.2 Forward Looking Consumers

Franco Modigliani and Milton Friedman argued that the key shortcoming of the Keynesian theory of consumption is the static nature of consumption decisions in this framework - consumption depends only on current incomes and not on future income or wealth. In practice, since households can borrow against future income (i.e. borrow today to raise consumption today and use future income to repay debts, therefore having lower consumption than income today), the decision to consume or save is fundamentally a dynamic one, where households choose consumption based not on current incomes but instead on the present value of all income they will receive over their lifetime. Recall that we define the Present Value of Lifetime Resources ($PVLR$) as

$$PVLR_t = \sum_{s=t}^{t+X-1} \frac{Y_s - T_s + Tr_s}{(1+r)^{s-t}}$$

In the permanent income theory of consumption, consumers choose their consumption and savings in each period to maximize the present discounted value of their utility. Suppose the planning horizon for a consumer is X years. For simplicity, suppose the real interest rate is constant at r. The problem consumers solve can be written as

$$\max_{\{C_t, C_{t+1}, \ldots, C_{t+X-1}, S_{t+1}, S_{t+2}, \ldots, S_{t+X}\}} U_t = \sum_{s=t}^{s=t+X-1} \frac{u(C_s)}{(1+\rho)^{s-t}}$$

subject to the sequence of budget constraints

$$C_s + S_{s+1} = (1+r)S_s + Y_s - T_s + Tr_s, \quad s = t, t+1, \ldots, t+X-1$$

and an initial value for S_t, which we'll just assume is zero (that is, we assume a consumer is born with zero wealth). In this problem,

- $u(\cdot)$ is a period utility function, sometimes called a felicity function, which captures the utility earned at each date from consumption at that same date. We will assume throughout that $u'(\cdot) > 0$ - consumers always prefer more consumption to less, so the marginal utility of consumption is strictly positive - and that $u''(\cdot) < 0$ - so marginal utility decreases with higher consumption.

- r is the real rate of return the consumer earns by saving.
- Y_s is the real income of the consumer at date s.
- T_s is the real value of taxes paid by the consumer and Tr_s is the real value of transfers received by the consumer.
- S_{s+1} is the amount the consumer saves at date s, on which she earns interest income at date $s+1$. The timing convention here, which uses the time subscript to denote the date at which the consumer receives income or spends on current consumption, is common in macroeconomics.

Notice that at date t the consumer is choosing the entire path for consumption and savings at all dates between t and $t+X-1$. This path is chosen based on the entire path for income, taxes and transfers that the consumer foresees between t and $t+X-1$. The consumer is thus *forward looking* - her choice of current consumption depends not just on current disposable income $Y_t - T_t + Tr_t$ but on the entire path of disposable income. Also notice that the consumer doesn't face one budget constraint here - she faces X budget constraints, one for each date. The budget constraint at date s says that the amount the consumer consumes C_s and the amount that she saves S_s must add up to her total resources, which are total income net of taxes and transfers $Y_s - T_s + Tr_s$ and the total amount the consumer had saved yesterday, on which she earns interest.

Let's now solve the household's problem. Before we write down a Lagrangian and take first order conditions, consider the household's budget constraints. At dates t, $t+1$ we know that

$$\begin{aligned}
C_t + S_{t+1} &= (1+r)S_t + Y_t - T_t + Tr_t \\
C_{t+1} + S_{t+2} &= (1+r)S_{t+1} + Y_{t+1} - T_{t+1} - Tr_{t+1} \\
C_{t+2} + S_{t+3} &= (1+r)S_{t+2} + Y_{t+2} - T_{t+2} - Tr_{t+2} \\
C_{t+3} + S_{t+4} &= (1+r)S_{t+3} + Y_{t+3} - T_{t+3} - Tr_{t+3} \\
&\vdots
\end{aligned}$$

Dividing the second equation by $1+r$ and adding the first two equations, we get

$$C_t + \frac{C_{t+1}}{1+r} + \frac{S_{t+2}}{1+r} = Y_t - T_t + Tr_t + \frac{Y_{t+1} - T_{t+1} - Tr_{t+1}}{1+r}$$

Divide the third equation by $(1+r)^2$ and add it to the equation above to get

$$C_t + \frac{C_{t+1}}{1+r} + \frac{C_{t+2}}{(1+r)^2} + \frac{S_{t+3}}{(1+r)^2} = Z_t + \frac{Z_{t+1}}{1+r} + \frac{Z_{t+2}}{(1+r)^2}$$

where we use the notation $Z_s \equiv Y_s - T_s + Tr_s$ for levity. We can continue this way until period $t+X-1$ to get

$$\sum_{s=t}^{s=t+X-1} \frac{C_s}{(1+r)^{s-t}} + \frac{S_{t+X}}{(1+r)^{t+X-1}} = \sum_{s=t}^{s=t+X-1} \frac{Y_s - T_s + Tr_s}{(1+r)^{s-t}}$$

Since the consumer does not care about consumption at any date after $t + X - 1$, it does not make sense for her to save anything at that date - so for her to be optimizing, she should set $S_{t+X} = 0$. Notice that the right hand side of the equation above is just the Present Value of Lifetime Resources for the consumer as defined in section 2 above! Thus, the large number of budget constraints reduces to just one lifetime budget constraint, of the form

$$\sum_{s=t}^{s=t+X-1} \frac{C_s}{(1+r)^{s-t}} = PVLR_t$$

Crucial for this result is that households have access to a frictionless financial market where they can save and invest any amount they choose at the same interest rate r.

We now solve the household's problem directly. Let λ be the Lagrange Multiplier on the household's problem. The household's Lagrangian is

$$\mathcal{L}_t = \sum_{s=t}^{t+X-1} \frac{u(C_s)}{(1+\rho)^{s-t}} + \lambda \left[PVLR_t - \sum_{s=t}^{s=t+X-1} \frac{C_s}{(1+r)^{s-t}} \right]$$

The first-order condition for consumption at date s is

$$\frac{u'(C_s)}{(1+\rho)^{s-t}} = \frac{\lambda}{(1+r)^{s-t}}$$

Dividing the FOC for consumption at date $s+1$ by the FOC for consumption at date s and rearranging, we get

$$\frac{u'(C_{s+1})}{u'(C_s)} = \frac{1+\rho}{1+r}$$

$$u'(C_s) = \frac{1+r}{1+\rho} u'(C_{s+1}) \tag{1}$$

This equation is called the Euler Equation, and is at the heart of how macroeconomists think about consumption behavior by households. To interpret the Euler equation, consider a household contemplating whether to save a dollar. A dollar saved reduces consumption today, which reduces the consumer's utility by the marginal utility of consumption - hence the left hand side represents the "cost" to the consumer, in utility units, of saving a dollar. However, a dollar saved today will earn interest, and so be worth $1 + r$ dollars tomorrow, raising consumption tomorrow by $1 + r$. Note that the impact of possible inflation between today and tomorrow is already captured by the fact that the interest rate we apply is the real interest rate. This increased consumption tomorrow will raise the consumer's utility by tomorrow's marginal utility $u'(C_{s+1})/(1 + \rho)$ where we need to divide by the $1 + \rho$ to account for discounting - a utility derived from consuming one unit tomorrow is worth only $1/(1 + \rho)$ times the utility that would be derived from consuming the same unit today.

To get a bit more intuition, let's consider the special case of isoelastic utility,

$$u(C_s) = \frac{C_s^{1-\frac{1}{\sigma}} - 1}{1 - \frac{1}{\sigma}}$$

where $0 < \sigma < \infty$ is the Intertemporal Elasticity of Substitution. Note that this implies that the marginal utility function is $u'(C_s) = C_s^{-1/\sigma}$. The Euler Equation is now

$$C_s^{-1/\sigma} = \frac{1+r}{1+\rho} C_{s+1}^{-1/\sigma} \implies 1 + g_{C,s} \equiv \frac{C_{s+1}}{C_s} = \left(\frac{1+r}{1+\rho}\right)^{\sigma}$$

Take logs on both sides of the Euler Equation. We have,

$$\log(1 + g_{C,s}) = \sigma \left(\log(1 + r_s) - \log(1 + \rho)\right)$$
$$\implies g_{C,s} \approx \sigma(r_s - \rho)$$

where the second line uses the approximation $\log(1 + x) \approx x$ for small x. This equation allows us to more easily interpret the intertemporal elasticity of substitution - it is just the change in the growth rate of consumption for a unit increase in the real interest rate. Empirically, σ is low for savers, who prefer a stable growth rate of consumption and respond to changes in interest rates with only small changes in consumption behavior. It is higher for borrowers, who tend to respond much more strongly to changes in the interest rates.

The Euler equation also allows us to interpret the fact that over most of recorded history, real interest rates have been positive. Rearranging the Euler equation gives us

$$r_s = \rho + \frac{g_{C,s}}{\sigma}$$

Over the long run, consumption growth has been relatively stable at around 2% a year, and ρ and σ are thought of as stable parameters governing preferences with positive values. The Euler Equation then predicts that the real interest rate must be positive to allow for this consumption growth to occur.

As an aside: How can we tell if the assumption of isoelastic utility makes sense? Note that one prediction of the model with isoelastic utility is that in the absence of any trend in real interest rates, there should be no trend in consumption growth rates - that is, if $1 + r_s \equiv 1 + r$ for all dates s then we must have $g_{C,s} = ((1 + r)/(1 + \rho))^{\sigma} \equiv g_C$ at all dates. The data do indeed suggest that over the 20th century, consumption growth has been relatively steady at about 2% a year and there is no trend in the real interest rate, implying that the data do not reject this model.

4.2.3 The Path of Consumption

To get a bit more intuition, it is worth solving explicitly for the entire path of consumption. Using the Euler Equation, we have

$$C_t = \left(\frac{1+r}{1+\rho}\right)^{-\sigma} C_{t+1} = \left(\frac{1+r}{1+\rho}\right)^{-2\sigma} C_{t+2} = ... = \left(\frac{1+r}{1+\rho}\right)^{\sigma(X-1)} C_{t+X-1}$$

These equations allow us to substitute for all the terms $C_s, s = t+1, t+2, \ldots, t+X-1$ in the intertemporal budget constraint, and express the equation in terms of C_t only. Recall that the budget constraint was

$$\sum_{s=t}^{s=t+X-1} \frac{C_s}{(1+r)^{s-t}} = PVLR_t$$

Substituting for C_s as above, we get

$$\begin{aligned}
\sum_{s=t}^{s=t+X-1} \frac{C_t}{(1+r)^{s-t}} \left(\frac{1+r}{1+\rho}\right)^{\sigma(s-t)} &= PVLR_t \\
\implies C_t \sum_{s=t}^{s=t+X-1} \left(\frac{(1+r)^{\sigma-1}}{(1+\rho)^{\sigma}}\right)^{s-t} &= PVLR_t \\
\implies C_t \left(\frac{1-\left(\frac{(1+r)^{\sigma-1}}{(1+\rho)^{\sigma}}\right)^X}{1-\frac{(1+r)^{\sigma-1}}{(1+\rho)^{\sigma}}}\right) &= PVLR_t \\
\implies C_t = PVLR_t \left(\frac{1-\frac{(1+r)^{\sigma-1}}{(1+\rho)^{\sigma}}}{1-\left(\frac{(1+r)^{\sigma-1}}{(1+\rho)^{\sigma}}\right)^X}\right)
\end{aligned}$$

It is clear that

- Any changes in the path of income between dates t and $t+X-1$ that leave $PVLR_t$ unchanged will lead to no changes in consumption at all.
- Consumption growth over time requires that $r \geq \rho$ - that is, the benefit of saving offsets the effect of discounting on utility.
- If $r = \rho$ then we have $C_t = C_{t+1}$ for all dates, so $C_t = \frac{PVLR_t}{X}$.
- A temporary change in income that affects Y_t, T_t or Tr_t will typically lead to small changes in $PVLR_t$. Thus, these changes should lead to small changes in current consumption C_t. Only permanent changes in income, which affect $PVLR_t$ substantially, can lead to changes in C_t.
- The fact that $PVLR_t$ summarizes the entire path of consumption going forward from today until the end of the household's planning horizon means that the impact of any changes in income that are **anticipated** by the household should be zero. For instance, students who expect to get a job shortly with a high salary should not respond to the anticipated increase in income in the future - rather, the fact that they will receive higher incomes in the future should already be reflected in their consumption today. Thus, only *unexpected* changes in the path of future incomes should affect current consumption.

4.2.4 Consumption Taxes and Consumption

Let's modify the framework here slightly to allow for consumption taxation. Suppose that households solve the problem

$$\max_{\{C_t, C_{t+1}, \ldots, C_{t+X-1}, S_{t+1}, S_{t+2}, \ldots, S_{t+X}\}} U_t = \sum_{s=t}^{s=t+X-1} \frac{u(C_s)}{(1+\rho)^{s-t}}$$

subject to the sequence of budget constraints

$$(1+\tau_{C,s})C_s + S_{s+1} = (1+r)S_s + Y_s - T_s + Tr_s, \quad s = t, t+1, \ldots, t+X-1$$

and an initial value for $S_t = 0$, with the isoelastic utility function

$$u(C_s) = \frac{C_s^{1-1/\sigma}}{1-1/\sigma}$$

We can proceed exactly as above (and the problem set asks you to show this!) to derive the following Euler equation for this framework.

$$\frac{C_s^{-1/\sigma}}{1+\tau_{C,s}} = \frac{1+r}{1+\rho} \frac{C_{s+1}^{-1/\sigma}}{1+\tau_{C,s+1}}$$

Rearranging terms,

$$\left(\frac{C_{s+1}}{C_s}\right)^{1/\sigma} = \frac{1+r}{1+\rho} \frac{1+\tau_{C,s}}{1+\tau_{C,s+1}}$$

Taking logs on both sides and using our log approximation $\log(1+x) \approx x$, we obtain

$$g_{C,s} = \sigma\left(r - \rho + \tau_{C,s} - \tau_{C,s+1}\right)$$

This equation shows that the effect of taxes on the growth rate of consumption depends on the change in the consumption tax rate between the two periods. A tax cut today financed by a tax cut tomorrow will, all else equal, reduce consumption growth. If we hold $PVLR$ fixed in this experiment, the fall in consumption growth will occur via a rise in consumption today (i.e. a fall in savings today) and a fall in consumption tomorrow (since tomorrow, there will be less interest income). As a result, changes in consumption tax rates can be a powerful tool to combat recessions, since sales tax cuts boost consumption in the present.

What are the underlying economics of this result? An increase in a consumption tax can be thought of as an increase in the price of consumption at the date in which it is levied. A cut in current consumption taxes relative to future consumption taxes reduces the relative price of consumption today as compared to consumption in the future, inducing the household to spend more on consumption in the present (when taxes, and the price of consumption are low) than the future (when taxes, and the price of consumption, are relatively high). This effect is called *intertemporal substitution* - forward looking households plan their consumption in such a way that the marginal value of a dollar spent on consumption is equated across all dates, accounting for the effect of taxes, interest rates and discounting.

4.2.5 The Evidence for Consumption Smoothing

The theory of consumption smoothing is predicated on some stark assumptions, the strongest of which is that households have access to perfectly competitive and flexible financial markets in which they can both borrow and save at a common interest rate r. This is admittedly a simplification - in the real world, consumers typically pay higher interest rates to borrow than they do to save. There are a number of reasons this can be the case:

- Since lending to households is risky, borrowers typically pay a risk premium to lenders to compensate them for the risk of default. Differential risk profiles of typical borrowers can explain why, for instance, some financial instruments such as credit cards have higher interest rates than typically safer lending, such as on mortgages.
- Financial institutions typically have market power, which allows them to charge higher interest rates on loans and pay lower interest rates on deposits than their cost of funding.

However, even in the presence of a wedge between deposit and lending rates, most households in advanced market economies like the US have access to financial markets, and therefore have the ability to smooth consumption by borrowing when income is low and saving when it is high. There is ample evidence that households do indeed engage in consumption smoothing.

- Households are exposed to business cycle fluctuations in their incomes, and yet maintain smoother profiles of consumption than their income over the business cycle (another way of saying this is that the saving rate is *procyclical* - households save more when their incomes are higher and dissave more when their incomes are lower).
- Savings rates follow income over the life-cycle - that is, most households save a higher and higher fraction of their income until their income peaks around the age of 40-55 years.
- Studies tend to estimate small MPCs out of changes in wealth caused by unexpected transitory shocks like unexpected capital gains on wealth held.
- Consumption of services is more stable than nondurables consumption, which in turn is more stable than the consumption of durables. Consumption of durables (which includes cars and heavy household appliances) typically behaves more like investment over the business cycle, since it involves infrequent purchases of items which provide service flows over long periods. Households should prefer to smooth the consumption of the services provided by durable goods, not the purchases of durable goods per se - that is, households probably want to have a working refrigerator in their house at all times to smooth their consumption of refrigeration services, but this does not mean that households need to purchase refrigerators at a consistent rate over time.

- Individuals with incomes that are more seasonal (including farmers, construction workers or workers receiving large year-end bonuses) do not have more seasonal fluctuations in consumption.

However, there is some evidence suggesting that the simplest theory of consumption smoothing is not a complete theory of consumption and saving behavior. For instance,

- Consumption is not perfectly flat over the lifecycle.
- In response to "unusual" changes in income that are temporary, such as stimulus checks, researchers tend to find relatively large MPCs, on the order of 0.4.
- The amount of savings that a permanent income consumer would choose given a typical income profile tends to produce too little wealth accumulation over time. Further, the theory predicts that households should choose to eventually run down their savings to 0, whereas in reality a large fraction of wealth accumulation comes from bequests across generations. While the theory can be easily modified to allow for a bequest motive for consumption, matching the data on wealth inequality usually requires assigning an outsized role to these bequests.

4.2.6 Savings rates across countries

Our simple theory can be used to shed some light on the wide range of savings rates seen in countries around the world. Explaining these differences requires us to consider:

- **Demographics:** Our theory suggests that if income follows a hump-shaped profile, savings rates will be low for the young and the old households. Thus, the savings rate of a country must be decreasing in the *dependency ratio*, the share of a country's population that follows outside the prime working ages of 15-65.
- **Tax rates:** In countries where consumption taxes are particularly high, households should choose to consume less and enjoy more leisure optimally, suggesting such societies should have high saving rates.
- **Liquidity constraints:** If households face trouble in accessing credit when their incomes are low, then households should choose to respond by saving more to ensure that they have a "buffer stock" to rely on when incomes are low.
- **Growth rates:** If households perceive robust income growth in the future, they will respond ex-ante by raising consumption, which should depress savings rates today.
- **Forced saving:** In some countries, policies regarding retirement benefits require all individuals to save certain fractions of their income at a minimum.
- **Safety net:** Countries with poor safety nets will have higher saving rates, as households try to build up savings they can rely on when incomes are lower.

4.3 Capital Demand and Investment

In our simple model, savings behavior generates the supply of capital for firms. We now turn to the demand for capital, which is generated by firms' optimal decisions of how much capital to utilize.

4.3.1 The user cost of capital

We model firms choosing how much capital K to rent and how much labor N to hire, given the price P of output Y and the interest rate r. We assume that once used in production, capital loses a fraction δ of its value, a phenomenon we call depreciation[43]. The firm produces output using the production function $Y = AF(K, N)$. It solves the problem

$$\max_{K,N} \Pi = [AF(K,N) - wN - \delta P_K K](1 - \tau_K) - rP_K K$$

where P_K is the price of capital and τ_K is the tax rate applied to profits. Note that as is conventional, taxes are applied on the firm's profits after accounting for

How do we interpret this problem? Ignore taxes. Consider an entrepreneur who is choosing between not doing anything, and receiving value 0, and

- borrowing P_K dollars at interest rate r to buy a unit of a machine,
- and tomorrow, selling the machine at a price $P_K(1 - \delta)$,
- and using the proceeds to repay the loan.

What is the *cost* of renting the machine via this transaction? Well, you pay an amount $(1 + r)P_K$ to the lender, but you can offset this with the amount $P_K(1 - \delta)$ you earned by selling the machine back, for a total net cost of $(1 + r)P_K - (1 - \delta)P_K = (r + \delta)P_K$. Clearly, you will be willing to engage in this transaction only if the extra unit of capital you got via this transaction produces at least $(r + \delta)P_K$ dollars of revenue. But we know that, by the definition of the marginal product of capital MPK, an extra unit of capital produces MPK units of output, which means $P \times MPK$ units of revenue. This suggests that the optimal rule for capital, which should equate the marginal benefit of the unit of capital with the cost associated with each unit of capital, should be of the form

$$P \cdot MPK = (r + \delta)P_K \implies MPK = (r + \delta)\frac{P_K}{P}$$

Let's use math to verify our intuition. The first order condition with respect to capital K is

$$(1 - \tau_K)\left(P \times \frac{dAF(K,N)}{dK} - \delta P_K\right) = rP_K \implies MPK = \left(\frac{r}{1 - \tau_K} + \delta\right)\frac{P_K}{P}$$

[43] Note that depreciation is a stand in for a wide variety of reasons why capital becomes less valuable over time, including physical wear and tear, the replacement of machine components and technological obsolescence.

which reduces exactly to the expression we expected if $\tau_K = 0$. The term on the right hand side of this expression is called the **user cost of capital**, and summarizes the total cost of investment for a firm. The firm's investment decision thus boils down to it choosing investment to hit a target capital stock, such that at this capital stock and the optimal labor hiring decision, the equation $MPK = uc$ holds. This is a powerful result, since it tells us that to characterize the effect of any change in the economy on the firm's investment decisions, we only need to think about how the change will affect the marginal product of capital and the user cost of capital. As we will see shortly, in a long-run equilibrium, the user cost of capital is usually responsive only to a rather small set of shocks.

For instance, consider a permanent shock to TFP that leaves the user cost of capital unchanged, and assume that production is Cobb-Douglas, implying that

$$Y = AK^{\alpha}N^{1-\alpha} \implies MPK \equiv \frac{dY}{dK} = \alpha AK^{\alpha-1}N^{1-\alpha} = \alpha\frac{Y}{K}$$

We know that there is no change in the long-run user cost by assumption. But the equation $MPK = uc$ then means that MPK is also unchanged. Since $MPK = \alpha Y/K$, the fixed MPK means that there must be no change in the long-run K/Y ratio either!

What's going on? On impact, in the short run (where K is fixed), the shock leads to an increase in output produced, and an immediate fall in K/Y as the denominator jumps but the numerator doesn't. The fall in K/Y leads to a rise in MPK and in the return to saving in capital. Households respond by saving more, leading to an increase in the capital stock over time. Eventually, the capital stock stabilizes at a higher level. Our argument above implies that in fact, it stabilizes at a level that restores the original K/Y ratio.

What's happening in the other important market, the labor market? On impact, the rise in TFP leads to a rise in the marginal product of labor, which would, all else equal, shift the labor demand curve to the right. In addition, households face an unexpected and permanent increase in their incomes, as the higher TFP should raise wages permanently. Since the change is permanent, households should respond by working less, leading to a leftward shift in the labor supply curve. The amount of labor input is ambiguous in the short run, but wages unambiguously rise.

Over time, as capital keeps rising, labor demand keeps rising as well. In the long run, the level of labor demand is permanently higher, and the K/N ratio is permanently higher as well. This in turn corresponds to higher real wages as well.

4.3.2 Investment and Capital

The capital stock at any date is a stock, that is measured at a given point of time. The change in the capital stock between two dates occurs as old capital loses value through depreciation and new investment creates new capital. We have

$$K_{t+1} = (1-\delta)K_t + I_t$$

Suppose investment I_t is fixed over time. Will capital settle down to a steady-state value at which it will remain constant? To answer this, we start by setting capital at the two dates equal. We have,

$$K = (1-\delta)K + I \implies I = \delta K \iff K = \frac{I}{\delta}$$

This equation says that the capital stock will converge to the level such that new investment is exactly enough to offset the depreciation of the capital stock at each date. We can also use this equation to answer the related, but very different question of what level of investment is required to keep the capital stock constant - investment should just equal the amount of depreciation that occurs each period.

We are finally ready to develop a theory of investment. In this model, profit maximizing firms first choose a target capital stock that is consistent with the profit maximization condition, $MPK = uc$. They then choose investment so that they can hit this target capital stock level next period. Thus, any force that raises the target capital stock tomorrow relative to today (or conversely, reduces today's capital stock relative to tomorrow's target capital stock) will raise investment. This includes forces that raise the expected MPK in the next period, including permanent increases in TFP, permanent increases in labor inputs or temporary shocks to the capital stock (such as disasters that destroy some capital). This also includes forces that reduce the user cost today (such as permanent reductions in the real interest rate, capital taxes or the relative price of capital P_K/P).

5 Combining the Markets: A Dynamic Macroeconomic Model

In topics 2 and 3, we have covered the basics of a modern macroeconomic model. To recap, these models have

- Two main sets of agents, **households** and **firms**.
- Households make **labor supply - leisure** decisions and **consumption - savings** decisions. These two decisions characterize labor supply and the supply of savings for investment purposes, as well as pinning down the path of consumption over time.
- Firms choose how much labor to hire and how much capital to rent from the household, their decisions pinning down the demand for labor and the demand for goods for investment purposes.

This section shows you how to combine the elements of the two topics together to obtain a complete general equilibrium model. We will start with a simple special case that can be solved almost entirely by hand. We will then lay out a complete dynamic model. Throughout this section, we will work with models that only have two periods, $t = 1, 2$. We will study infinite horizon models later.

5.1 A Dynamic Macroeconomic Model: Getting Started

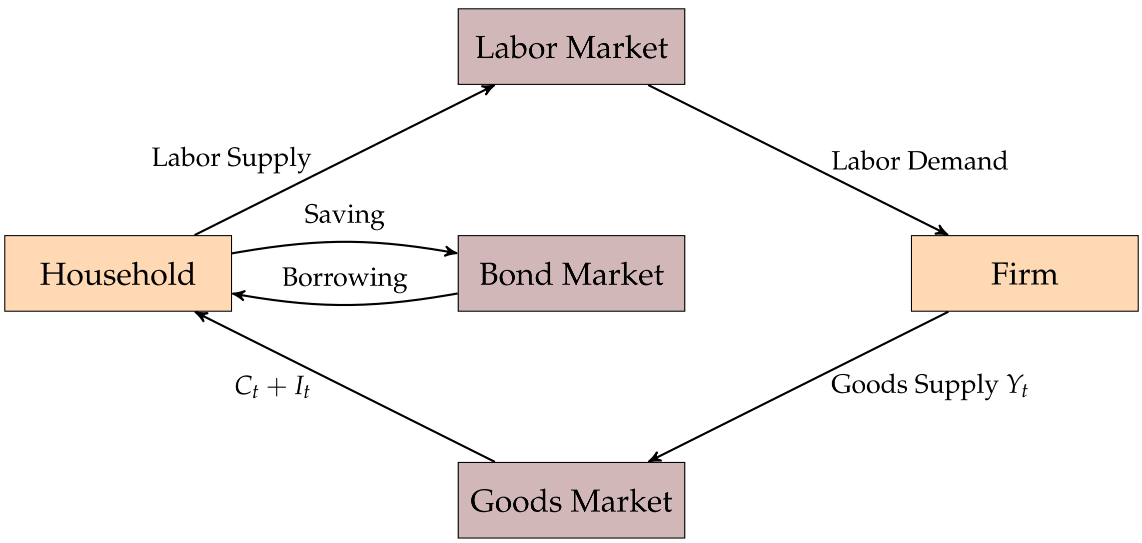

Figure 1: A Simple Macroeconomic Model: Agents and Flows

In this simplest model, there are two agents - households and firms - who interact in three markets, the markets for labor, goods and bonds.

- In the market for **labor**, households supply labor and are paid wages, and firms demand labor which they use in the production of output. For this simple model, we assume that
  - Labor is the only factor of production, and that output is produced using the production function
  $$Y_t = A_t N_t, t = 1, 2$$
  - Labor supply is *perfectly inelastic* - all consumers are endowed with 1 unit of time, and they are willing to supply this 1 unit of time to the labor market irrespective of the wage rate.

  Both of these assumptions will be relaxed when we do the complete model in the next section.

- In the market for **goods**, firms produce output Y_t. We will for simplicity treat output as the *numeraire* in this economy, which means we define prices in units of final goods. This immediately means that the price of final goods is 1, since the number of final goods equivalent to one unit of final goods is, well, one unit of final goods. When we say the wage rate is w, we mean that one hour of labor leads to the consumer receiving w units of final goods worth of compensation. In the goods market, final goods will be sold to households, who will then choose consumption C_t and savings S_t at the two dates.

- In the market for **bonds**, households interact with each other. We will assume that households enter date 1 with no debt outstanding or savings. At date $t = 1$, a household choosing to *borrow* will *issue* bonds (i.e. try to sell bonds), and a household choosing to *save* will *buy* bonds from the borrowers. A household purchasing a bond worth \$1 from the borrower at date $t = 1$ is promised a repayment of $\$1 + r_2$ at date $t = 2$. We assume that the bond market is only open at date $t = 1$. This isn't a huge assumption. Since the world ends at the end of date 2, no agent who buys a bond at date 2 will ever be repaid, which means they lose the entire amount they spend on buying bonds - and would have been far better off just consuming that amount at date 2 instead.

5.1.1 What this Model will take as Exogenous and Endogenous

Our model will take productivity in the two periods A_1, A_2 as exogenous, and normalize the total population to 1. The model's **solution** is a set of values for consumption, savings and labor supply at each date - which makes these the endogenous variables. Let's characterize the household and firm problems more carefully now.

5.1.2 Households

Households solve the utility maximization problem defined as

$$\max_{C_1,C_2,S_2} \frac{C_1^{1-1/\sigma}}{1-1/\sigma} + \frac{1}{1+\rho}\left[\frac{C_2^{1-1/\sigma}}{1-1/\sigma}\right]$$

$$\text{subject to}$$

$$C_1 + S_2 = w_1$$

$$C_2 = w_2 + (1+r_2)S_2$$

where w_1, w_2 are the wage rates at the two dates and r_2 is the interest rate in the bond market. Households take these prices as given when they decide how much to work and how much to save. Notice that households have isoelastic utility over consumption in the two periods, and also have an isoelastic disutility of working. Further, the household's preferences are *time-separable* since we can split the total lifetime discounted utility into the sum of utility at date $t = 1$ and discounted utility at date $t = 2$.

Before we take first order conditions, it's worth combining the budget constraints together and eliminating saving S_2 from the problem, by dividing the second constraint by $1 + r_2$ and adding the two constraints to get the equivalent problem

$$\max_{C_1,C_2} \frac{C_1^{1-1/\sigma}}{1-1/\sigma} + \frac{1}{1+\rho}\left[\frac{C_2^{1-1/\sigma}}{1-1/\sigma}\right]$$

$$\text{subject to}$$

$$C_1 + \frac{C_2}{1+r_2} = w_1 + \frac{w_2}{1+r_2}$$

Let λ be the Lagrange Multiplier on the constraint, so that the Lagrangian is

$$\mathcal{L} = \frac{C_1^{1-1/\sigma}}{1-1/\sigma} + \frac{1}{1+\rho}\left[\frac{C_2^{1-1/\sigma}}{1-1/\sigma}\right] + \lambda\left[w_1 + \frac{w_2}{1+r_2} - C_1 - \frac{C_2}{1+r_2}\right]$$

The first order conditions of the problem with respect to C_1, C_2 give us

$$C_1^{-1/\sigma} = \lambda$$

$$\frac{C_2^{-1/\sigma}}{1+\rho} = \frac{\lambda}{1+r_2}$$

and the first order condition with respect to λ is just the budget constraint. Combining these equations to eliminate the Lagrange Multiplier and adding the budget constraint, we obtain two equations in C_1, C_2:

$$C_1^{-1/\sigma} = \frac{1+r_2}{1+\rho} C_2^{-1/\sigma} \tag{2}$$

$$C_1 + \frac{C_2}{1+r_2} = w_1 N_1 + \frac{w_2 N_2}{1+r_2} \tag{3}$$

The first equation is just the Euler equation, and the second is the present value budget constraint.

5.1.3 The Firm

The representative firm solves a static problem of deciding how much labor to hire. It solves

$$\max_{N_t}(A_t - w_t)N_t$$

for each date $t = 1, 2$, taking the wage rate w_t as given. The firm's first order condition with respect to N_t just sets $A_t - w_t = 0$. To interpret this, observe that under the production function $Y_t = A_t N_t$ there are no diminishing returns to labor - each unit of labor hired always produces A_t units of output and thus produces net profits $A_t - w_t$. If $A_t > w_t$ it is optimal for the firm to hire an unlimited amount of labor, while if $A_t < w_t$ the firm's optimal hiring decision is to not hire anyone. If $A_t = w_t$, the firm is indifferent about the number of hours it hires.

5.1.4 Market Clearing

Market clearing conditions are at the heart of modern macroeconomics. At the most basic level, a market clearing condition is an equilibrium condition that simply requires that prices and allocations be such that the total demand and total supply of a given good or factor of production are equal to each other. In our model, we have the following conditions.

Goods Market Clearing requires that the demand for goods, C_t, equals the total output of goods Y_t at each date $t = 1, 2$. Note that our model does not have investment since we have abstracted from capital for now, and we consider only closed economies without governments here.

Labor Market Clearing requires that the demand for labor from firms at the market wage rate w_t equals the supply of labor from households, which we fixed inelastically at 1.

Bond Market Clearing requires that total savings in bonds equals total borrowing in bonds at date $t = 1$.

5.1.5 An Equilibrium

Given a path $\{A_1, A_2\}$ for productivity, a **Dynamic General Equilibrium** of the simple macroeconomic model we have set up consists of

- An **Allocation** of consumption, savings, labor inputs

$$\{C_t, S_t, N_t\}_{t=1,2}$$

- A **path for wages** $\{w_t\}_{t=1,2}$ and a real interest rate in the bond market r_2

such that

- The household's choices of consumption and savings maximize the present value of its utility, i.e. equations 2,3 hold.
- The firm's choices of capital and labor at each date maximize its profits period-by-period.
- All markets clear.

5.1.6 Characterizing Equilibrium

It turns out that in this simple economy, we can solve for most of the endogenous quantities as functions of exogenous variables and parameters directly.

- **Start with the Labor Market.** Since labor supply equals 1 irrespective of the wage rate, we know that the equilibrium quantity of labor input will be 1. But we know that labor demand by firms is 0 for any $w_t > A_t$, infinite for $w_t < A_t$ and indeterminate for $w_t = A_t$. The *only* way for the labor market to clear is for $w_t = A_t$ and the quantity of labor to be pinned down by the labor supply condition. Thus, we immediately know that

$$w_1 = A_1, w_2 = A_2, N_1 = N_2 = 1$$

- **Move to the Goods Market Clearing Conditions.** We know that at each date $Y_t = C_t$. But we know from above that $N_1 = N_2 = 1$, so by the production function we know that $Y_1 = A_1, Y_2 = A_2$. Thus, we immediately know that $C_1 = A_t, C_2 = A_2$. We also know that $S_1 = S_2 = 0$.

  The fact that there are no savings in this economy might seem strange, but this is just a consequence of the fact that *in the aggregate*, bonds are in *zero net supply* - that is, for every household selling a bond in order to borrow there must be a household buying the bond in order to save, so *on net* there is no saving. Note that since all individuals in this model are identical, it cannot be that some agents would strictly prefer to save and some would strictly prefer to borrow in bonds - if any one agent strictly prefers to save than borrow, they must all strictly prefer to save. However, it is impossible for all agents to save or for all agents to borrow in bonds, since the only way to save in bonds is for there to be someone willing to sell them. The only way for the bond market to clear, then, is for consumers to be indifferent between borrowing and saving at the interest rate prevailing in the bond market - which is, in fact, exactly the condition implied by the Euler Equation.

- **Use the Euler Equation.** There's only one last thing we need to pin down: the real interest rate in the date 2 bond market, r_2. Rearranging the Euler equation gives
$$1 + r_2 = (1 + \rho)\left(\frac{C_2}{C_1}\right)^{1/\sigma}$$
Define g_A by the equation $1 + g_A = A_2/A_1$. Substituting $C_t = A_t$ and taking logs,
$$\log(1 + r_2) = \log(1 + \rho) + \frac{1}{\sigma}(\log(1 + g_A))$$
$$r_2 = \rho + \frac{g_A}{\sigma}$$
where the final line uses the approximation $\log(1 + x) \approx x$ and defines $1 + g_A = A_2/A_1$.

This completes our characterization of the model - we have solved for all the endogenous variables as functions of the exogenous variables and parameters only! This version of the model makes assumptions that ensure that we can solve for everything by hand. We now relax some assumptions and see what the implications of this are.

5.2 The Complete Two-Period Dynamic Macroeconomic Model

We now lay out a complete macroeconomic model and define *equilibrium* in the model. This model will retain most of the insights obtained from the last section - except, we will add capital back to the model. We will see that the equilibrium can be characterized as the solution to a set of equations. The next sections will continue to build on this simple model to study classic issues in macroeconomics, including growth, fiscal policy and monetary policy.

5.2.1 What this Model will take as Exogenous and Endogenous

Our model will take the following things as exogenous.

- The household's initial wealth S_0, which we will assume is K. In this economy, households will only be able to save by accumulating capital.
- Productivity in the two periods A_1, A_2.
- The prices of output, P, which we will normalize to 1. We will also assume that $P_K = 1$, that is, capital and consumption goods have the same price.

The model's **solution** is a set of values for

- Consumption, savings and labor supply at each date
- Investment and the capital stock at each date.

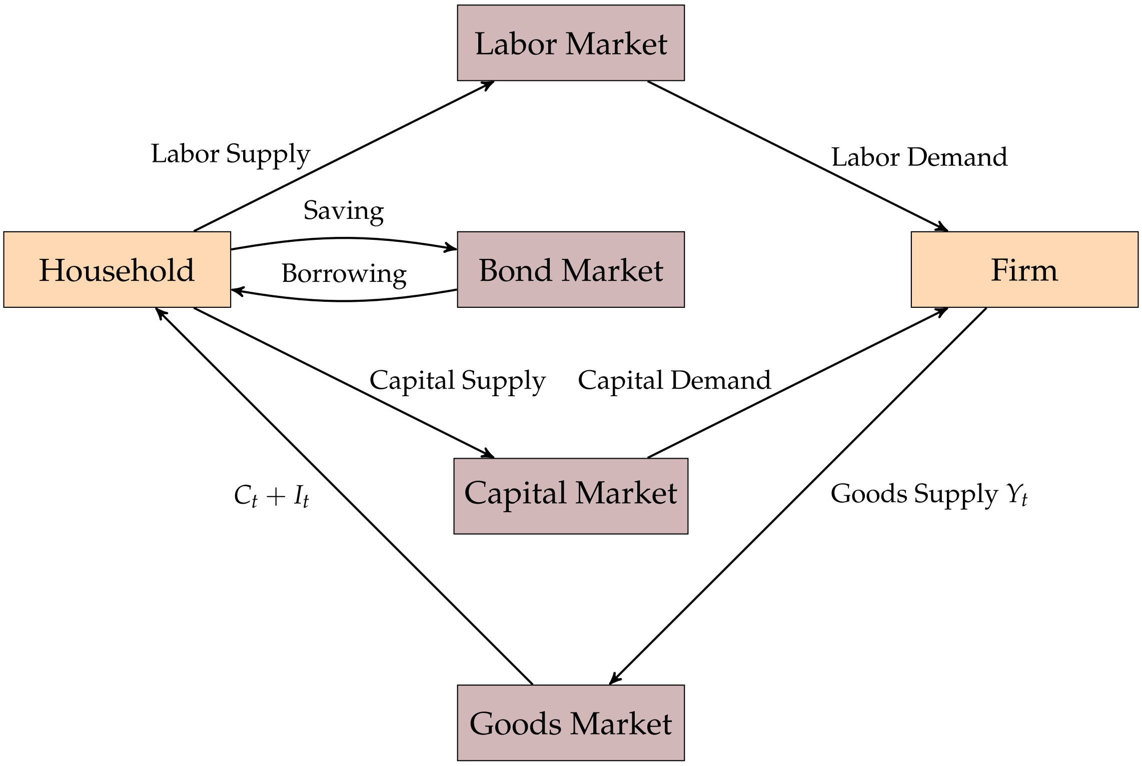

Figure 2: A Simple Macroeconomic Model: Agents and Flows

5.2.2 Households

A household is born with initial wealth $S_1 = K$. It lives for two periods, say 1 and 2. At each date, it chooses

- How much to consume and save C_t, S_t for $t = 1, 2$.
- How much to work, N_t for $t = 1, 2$.

When making decisions, the household takes wages w_t and interest rates r_t as given. That is, it does not take into account the impact of its own decisions on the prices in the economy. The household's problem can be written as

$$\max_{C_1, C_2, S_2, S_3, N_1, N_2} \frac{C_1^{1-1/\sigma}}{1-1/\sigma} - \frac{N_1^{1+1/\theta}}{1+1/\theta} + \frac{1}{1+\rho}\left[\frac{C_2^{1-1/\sigma}}{1-1/\sigma} - \frac{N_2^{1+1/\theta}}{1+1/\theta}\right]$$

$$\begin{aligned}
&\text{subject to} \\
&C_1 + S_2 = (1+r_1)S_1 + w_1 N_1 \\
&C_2 + S_3 = (1+r_2)S_2 + w_2 N_2 \\
&S_3 \geq 0 \\
&S_1 = K \text{ given}
\end{aligned}$$

Note that in this problem, we fix the household's initial wealth K and also prevent the household from choosing a negative level of saving in the final period (i.e. we don't allow the household to borrow in the final period, knowing full well that if it could, it would borrow an infinite amount in this period since there's no period left in which it has to repay!) Observe that under this constraint, it can never be optimal to save in period 3. If the household was saving \$1 in the final period, it could instead choose to consume that \$1, which would strictly raise its lifetime utility! Thus, we can ignore this constraint and simply set $S_3 = 0$.

After this, we combine the two constraints into 1, by eliminating S_2. Dividing the second constraint by $1 + r$ and adding the resulting equation to the first constraint, we obtain the equivalent problem

$$\begin{aligned}
\max_{C_1, C_2, N_1, N_2} & \frac{C_1^{1-1/\sigma}}{1-1/\sigma} - \frac{N_1^{1+1/\theta}}{1+1/\theta} + \frac{1}{1+\rho}\left[\frac{C_2^{1-1/\sigma}}{1-1/\sigma} - \frac{N_2^{1+1/\theta}}{1+1/\theta}\right] \\
& \text{subject to} \\
& C_1 + \frac{C_2}{1+r_2} = (1+r_1)K + w_1 N_1 + \frac{w_2 N_2}{1+r_2}
\end{aligned}$$

Let λ be the Lagrange Multiplier on the constraint, so that the Lagrangian is

$$\begin{aligned}
\mathcal{L} = \frac{C_1^{1-1/\sigma}}{1-1/\sigma} - \frac{N_1^{1+1/\theta}}{1+1/\theta} + \frac{1}{1+\rho}\left[\frac{C_2^{1-1/\sigma}}{1-1/\sigma} - \frac{N_2^{1+1/\theta}}{1+1/\theta}\right] \\
+ \lambda\left[(1+r_1)K + w_1 N_1 + \frac{w_2 N_2}{1+r_2} - C_1 - \frac{C_2}{1+r_2}\right]
\end{aligned}$$

The first order conditions of the problem with respect to C_1, C_2, N_1, N_2 give us

$$\begin{aligned}
C_1^{-1/\sigma} &= \lambda \\
\frac{C_2^{-1/\sigma}}{1+\rho} &= \frac{\lambda}{1+r_2} \\
N_1^{1/\theta} &= w_1 \lambda \\
\frac{N_2^{1/\theta}}{1+\rho} &= \frac{w_2 \lambda}{1+r_2}
\end{aligned}$$

Combining these equations to eliminate the Lagrange Multiplier and adding the budget constraint, we obtain four equations in the four variables C_1, C_2, N_1, N_2:

$$N_1^{1/\theta} = w_1 C_1^{-1/\sigma} \tag{4}$$

$$N_2^{1/\theta} = w_2 C_2^{-1/\sigma} \tag{5}$$

$$C_1^{-1/\sigma} = \frac{1+r_2}{1+\rho} C_2^{-1/\sigma} \tag{6}$$

$$C_1 + \frac{C_2}{1+r_2} = (1+r_1)K + w_1 N_1 + \frac{w_2 N_2}{1+r_2} \tag{7}$$

5.2.3 Firms

Firms rent capital at interest rate $r_t, t = 1, 2$ and face depreciation δ. They also hire labor. The problem they solve is static - at each date $t = 1, 2$, they solve the problem

$$\max_{K_t, N_t} A_t K_t^{\alpha} N_t^{1-\alpha} - w_t N_t - (r_t + \delta) K_t$$

The four first order conditions that characterize the path of labor and capital demand by the firm are

$$w_1 = (1-\alpha) A_1 \left(\frac{K_1}{N_1}\right)^{\alpha} = (1-\alpha)\frac{Y_1}{N_1} \tag{8}$$

$$w_2 = (1-\alpha) A_2 \left(\frac{K_2}{N_2}\right)^{\alpha} = (1-\alpha)\frac{Y_2}{N_2} \tag{9}$$

$$r_1 + \delta = \alpha A_1 \left(\frac{K_1}{N_1}\right)^{\alpha-1} = \alpha\frac{Y_1}{K_1} \tag{10}$$

$$r_2 + \delta = \alpha A_2 \left(\frac{K_2}{N_2}\right)^{\alpha-1} = \alpha\frac{Y_2}{K_2} \tag{11}$$

Given capital demand, we know that investment at date 1 must satisfy

$$I_1 = K_2 - (1-\delta) K_1$$

5.2.4 Market Clearing

In this model, we have four markets.

Bond Market Clearing requires that savings in bonds equals the total amount borrowed in bonds. As in the simple model, in the aggregate, savings by agents in bonds cancels out exactly with borrowing by agents in bonds, and thus the net aggregate savings of households in the bond market is zero.

Labor Market Clearing requires that labor demand equal labor supply,

$$N_1^D = N_1^S \quad ; \quad N_2^D = N_2^S$$

Capital Market Clearing requires that capital demand equal capital supply,

$$K_1^D = K_1^S \quad ; \quad K_2^D = K_2^S$$

An alternative formulation of these is that the wage rate that households take as given at each date when choosing labor supply must be the same as the wage rate that firms take as given when choosing labor demand. This is the formulation we will use below to solve the model. An analogous condition holds for capital market clearing, but we need to be a bit more careful when we characterize capital supply at the two dates. Note that capital supply at date 1 is equal to K, which is exogenous - it just equals the initial capital stock pre-determined somewhere else. Recall that investment today cannot alter the capital stock contemporaneously. By constrast, capital supply at date 2 is determined *endogenously* by savings choices made by households - it equals

$$K_2^S = (1-\delta)K_1^S + S_1$$

Finally, **Goods market clearing** just requires that at all dates, output is either consumed or invested. This implies that

$$Y_1 = A_1 K_1^{\alpha} N_1^{1-\alpha} = C_1 + I_1 = C_1 + K_2 - (1-\delta)K_1$$

5.2.5 An Equilibrium

Given initial capital K and a path $\{A_1, A_2\}$ for productivity, a **Dynamic General Equilibrium** of the simple macroeconomic model we have set up consists of

- An **Allocation** of consumption, savings, labor supplies and capital stocks

$$\{C_t, S_t, N_t, K_t\}_{t=1,2}$$

- A **path for the prices** $\{w_t, r_t\}_{t=1,2}$

such that

- The household's choices of consumption, savings and labor supply maximize the present value of its utility, i.e. equations 4,5,6,7 hold.
- The firm's choices of capital and labor at each date maximize its profits period-by-period, i.e. equation 8,9,10,11 hold.
- All markets clear.

5.2.6 Solving for the Equilibrium

Note that given exogenous values A_1, A_2, K, the dynamic model can be written as a system of nonlinear equations in the endogenous variables $\{C_1, C_2, N_1, N_2, K_2, w_1, w_2, r_1, r_2\}$.

This system of equations is

$$
\begin{aligned}
N_1^{1/\theta} &= w_1 C_1^{-1/\sigma} \\
N_2^{1/\theta} &= w_2 C_2^{-1/\sigma} \\
C_1^{-1/\sigma} &= \frac{1+r_2}{1+\rho} C_2^{-1/\sigma} \\
C_1 + \frac{C_2}{1+r_2} &= (1+r_1)K + w_1 N_1 + \frac{w_2 N_2}{1+r_2} \\
w_1 &= (1-\alpha) A_1 \left(\frac{K}{N_1} \right)^{\alpha} \\
w_2 &= (1-\alpha) A_2 \left(\frac{K_2}{N_2} \right)^{\alpha} \\
r_1 + \delta &= \alpha A_1 \left(\frac{K}{N_1} \right)^{\alpha-1} \\
r_2 + \delta &= \alpha A_2 \left(\frac{K_2}{N_2} \right)^{\alpha-1} \\
K_2 &= (1-\delta)K + A_1 K^{\alpha} N_1^{1-\alpha} - C_1
\end{aligned}
$$

The first 4 equations are the household's FOCs and the next 4 are the firm's FOCs. The final equation is just the law of motion for capital. These equations can be solved using a computer to obtain different values for the endogenous variables, as we change the exogenous variables. The way to do this is as follows.

- Note that we take exogenous variables $\{K, A_1, A_2\}$ and parameters $\{\theta, \sigma, \rho, \alpha, \delta\}$ as given and the endogenous variables $\{C_1, C_2, N_1, N_2, w_1, w_2, r_1, r_2, K_2\}$ are what we want to calculate.
- In your favorite programming language, start by defining constants with values equal to the exogenous variables and parameters.
- Write a subroutine (in MATLAB, this would be a separate `m` file for this function, and in Python, this would be a function defined the usual way using `def`) that takes the endogenous variables as inputs. The subroutine should calculate the differences between the left and the right hand sides of the equations above evaluated at and return this in a vector.

  In Python, this subroutine might look something like

$$
\begin{aligned}
&\texttt{def resid}(C_1, C_2, N_1, N_2, w_1, w_2, r_1, r_2, K_2)\texttt{:} \\
&\quad \texttt{res} \;=\; \texttt{np.zeros(9)} \\
&\quad \texttt{res[0]} = N_1^{1/\theta} - w_1 C_1^{-1/\sigma} \\
&\quad \texttt{res[1]} = N_2^{1/\theta} - w_2 C_2^{-1/\sigma} \\
&\quad \vdots \\
&\quad \texttt{res[8]} = K_2 + C_1 - (1-\delta)K - A_1 K^{\alpha} N_1^{1-\alpha} \\
&\quad \texttt{return res}
\end{aligned}
$$

- Use a numerical solver in the programming language to find a zero of this subroutine. Observe that at the set of values for the endogenous variables that sets the value of the subroutine to 0, the equations must all hold since that's how we defined the output of the subroutine! In both MATLAB and Python, you can use a routine called `fsolve` to find the zero. These routines typically require an initial guess for the solution, and unfortunately this is not always easy to guess.

6 Fiscal Policy

Government spending is a large fraction of GDP in most advanced economies. Changes in the timing and level of government spending and the taxes and public debt accumulation required to finance it have important effects on the economy, and as we'll see, are an invaluable policy tool for governments seeking to ameliorate the effects of recessions. In this section, we will explore some of these effects.

6.1 Government Spending

6.1.1 Government Consumption vs Government Investment

Economists sometimes distinguish between Government Consumption (which we will denote G^C) and Government Investment (which we will denote G^I). **Government consumption** should be thought of as public expenditure on goods and services delivered to the public contemporaneously, which add to current expenditures but do not raise the private or public capital stock. Examples include the provision of public goods like personnel for police or public administration services. **Government Investment**, by contrast, is thought of as public expenditure on the creation of new capital, such as infrastructure spending or education spending (which leads to the accumulation of *human* capital).

6.1.2 The Effects of Government Consumption

The key to understanding the effect of an increase in consumption G^C is to note that this increase is not costless - the government can only raise G by raising taxes or reducing transfers, either in the present or in the future.

Suppose the government raises G^C *permanently*, and finances this by reducing lump-sum transfers *permanently*, keeping tax rates unchanged. This must reduce the present value of lifetime resources for the household by an equivalent amount. As a result of this reduction, in the short run, households will tend to consume less and also work harder due to the income effect. If the increase in G^C does not affect the long-run level of TFP or the growth rate of output, it will leave the user cost of capital unchanged in the long run, which means firms will target the same level of the capital-labor ratio in the long run. Since labor supply has risen, firms will respond by investing more, and this will raise capital until the K/N ratio returns to its original level. In the long run equilibrium, there is no impact on wages, the capital-labor or capital-output ratios, or on the marginal product of labor.

Since households are consuming less and working more, it is tempting to argue that households must be worse off due to the increase in G^C. However, remember that households may value G^C directly in their utility function - that is, households may derive utility from living in safe neighborhoods with functioning public transport systems. What matters for evaluating the welfare gains from consumption G^C is whether the marginal increase in utility from a dollar spent on government consumption outweighs the marginal decrease in utility due to the income effects we just discussed.

6.1.3 The Effects of Government Investment

A simple way to model the effects of investment G^I is to assume that higher G^I raises productivity directly, so that A is higher. Consider a permanent increase in investment G^I financed by a decrease in transfers, keeping tax rates unchanged.

As a baseline, suppose the government chooses investment projects poorly, and that the benefits of G^I never materialize - that is, A is not higher. Then the increase in G^I is exactly like an increase in G^C, and has the same income effects, corresponding to lower consumption and less leisure. If investment G^I does not enter into utility directly - say because no one can use a bridge to nowhere - then the impact on welfare is unambiguously negative.

Suppose, however, that the investments materialize in the form of a much higher level of TFP. In the short run, if the increase in TFP is large enough, the implied increase in productivity can actually raise $PVLR$ by more than the cost of the investments. The income effects now go the other way, with higher consumption, investment and output. The impact on labor is ambiguous: on the one hand, the higher $PVLR$ will induce households to work less via the income effect, and on the other hand, the increase in A will raise labor productivity and wages, inducing more effort via a substitution effect. If the income effect dominates, we could see a decline in labor.

If the long-run level of TFP is higher but there is no impact on the growth rate of TFP, the long-run user cost of capital will not change, and the K/Y ratio will not change either. The higher level of A then implies that K/N and Y/N are both higher.

For context, the U.S. currently spends about 1.5% of GDP on infrastructure, down from 2% or so in earlier decades, which has led to a strong political push for raising infrastructure spending. Our discussion above shows that a falling infrastructure share of GDP and government spending is not necessarily a bad thing, since it all depends on the productivity boost we gain from the extra infrastructure built. Most economists think the U.S. should spend more currently than it does, but this assumes that infrastructure spending is well-prioritized. In practice, political forces can result in bad projects being funded, with politicians redirecting infrastructure projects to benefit their own constituents[44]. An independent body to prioritize infrastructure projects, say based on the model of the Federal Reserve Board, could mitigate some of these concerns.

[44] Sometimes referred to as "pork-barrel" spending in the financial press.

6.2 Tax Policy

6.2.1 Marginal and Average Tax Rates

Let y denote a household's pre-tax income and $T(y)$ denote the amount of taxes paid by the household as a function of its income. The burden of taxation on the household is usually summarized by two metrics:

- **Marginal Tax Rates**, the increase in taxes paid by a household for a marginal change in income. Mathematically, the marginal tax rate is the derivative of the tax function with respect to income, $T'(y)$.
- **Average Tax Rates**, the ratio of total taxes paid to total income, $T(y)/y$.

If tax rates are progressive, i.e. the marginal tax rate increases with income, then the average tax rate will in general be lower than the marginal tax rate.

To see a simple example that describes the difference between average and marginal tax rates, consider a heavily simplified version of the US tax system. Let D denote the level of income that is *deductible* for the purposes of taxation, i.e. given income y, the amount of *taxable income* is 0 if $y \leq D$ and $y - D$ if $y > D$. The tax function is given by

$$T(y) = \begin{cases} 0 & y \leq D \\ \tau(y - D) & y > D \end{cases}$$

where $\tau > 0$. Note that the marginal tax rate is 0 for $y \leq D$ and τ for $y > D$, while the average tax rate is 0 for $y \leq D$ and $\tau\left(\frac{y-D}{y}\right) = \tau - \tau D/y < \tau$ for $y > D$.

In practice, most countries have more than two tax brackets and offer a wide range of deductions to incentivize certain kinds of transactions.

6.2.2 Static and Dynamic Scoring

To study changes in tax revenues that occur when tax policies are changed, policymakers can conduct two kinds of analyses.

- **Static Scoring** is a method of calculating changes in tax revenues that arise purely from changes in the parameters of the tax function, assuming no changes in any choices made by households (including in labor supply, saving or investment). An advantage of static scoring is that it is simple and relatively model-free, but a significant drawback is that it ignores all responses by economic agents to changes in tax policy. Static scoring is used by policy organizations, including the non-partisan Congressional Budget Office, to determine the revenue impacts of tax changes.

- **Dynamic Scoring** involves taking into account changes in behavior (particularly with respect to labor supply and saving/investment behavior). While dynamic scoring provides a more complete answer that takes into account responses of agents to tax changes, it is much more complex to perform and requires assumptions about the nature of responses by individuals to tax changes.

To see an example of these different methods in action, consider the simple model of labor supply we studied in section 2. Suppose a government cuts the tax rate on labor income today, financing this by reducing G in the future.

- Under static scoring, we hold the number of hours worked and the (pre-tax) wage rate fixed. The result of cutting the tax rate today, keeping pre-tax income fixed, is clearly a reduction in tax revenue.
- Under dynamic scoring, we take into account the fact that the tax cut today will encourage an increase in labor supply, and the decrease in total tax burden will raise $PVLR$ and therefore tend to reduce labor supply. If the latter income effect dominates the first substitution effect, the tax cut will induce a fall in labor input that will further reduce the tax base as well, cutting tax revenue by even more than the static scoring exercise suggests.

6.2.3 Tax Reforms

A significant area of research in economics and an active area of debate in policy is the optimal design of tax policy, which seeks to minimize the efficiency losses associated with taxes subject to raising a given level of revenue. This debate is important in the US in particular, where

A policy proposal that is popular is a "flat tax", which would replace the current system with a single marginal tax rate and a single deduction. This system would reduce compliance costs (think of the time and resources devoted to producing tax forms, filing tax returns and verifying the information on these) and reduce distortionary effects on labor supply, potentially leading to increases in hours worked and consumption that are efficient. Further, implementing a flat tax would eliminate politically motivated tax breaks that distort economic incentives and lead to inefficient outcomes[45].

However, flat taxes also have their downsides. First, flat taxes would be regressive by design, since they would lead to higher post-tax and consumption inequality. This can, all else equal, reduce average welfare ex-ante. Further, they would eliminate tax breaks that might be justified on efficiency grounds, such as those designed to create economic incentives for activities associated with positive externalities (such as human capital accumulation or research and development). Finally, flat taxes and the resulting simplification of the tax code could reduce opposition to tax collection and scrutiny of

[45]For instance, in the US, the mortgage interest tax deduction allows buyers of houses to deduct interest paid on mortgages up to $500,000 in mortgage debt. While ostensibly designed to help first time home buyers, the program generates incentives for home buyers to use as much mortgage debt as possible when buying a house, and to buy houses that are larger than might be optimal for a given buyer.

government spending financed this way, therefore potentially allowing for increases in wasteful government spending.

6.2.4 Supply-Side Economics

Recall that changes in tax rates have competing substitution and income effects:

- A tax cut has a substitution effect that induces an increase in labor supply, since the after-tax return to working rises, all else equal.
- A permanent tax cut has an income effect as well - all else equal, a permanent tax cut is like a permanent increase in wages, which raises $PVLR$ and should tend to reduce labor supply.

Supply-side economics is a school of thought in public economics emphasizing the substitution effects of tax policies, arguing that if substitution effects are dominant, then lower marginal tax rates on labor income will raise labor supply.

A corollary of the approach taken by supply-side economists is the idea of the **Laffer Curve**, proposed by Arthur Laffer. The Laffer Curve is a plot of total tax revenue $T(y)$ against the marginal tax rate τ. Starting at very low levels of the marginal tax rate τ, increases in the marginal tax rate should have small disincentive effects on labor supply, so a small increase in the marginal tax rate should raise total tax revenue - as a result, for low τ, the Laffer curve is upward-sloping. Intuitively, in this range, the reduction in labor supply (or the shrinking of the *tax base*) is more than offset by the increase in the tax rate. However, as τ grows higher, the disincentive effects of higher marginal tax rates grow stronger, so the marginal increase in total tax revenue for a marginal increase in the tax rate falls. For high marginal tax rates, further increases in the tax rate lead to reductions in the tax base large enough that total tax revenue actually falls. The Laffer curve is thus downward sloping for high enough τ.

The values of the marginal tax rate τ for which total tax revenue is decreasing are said to be on the wrong side of the Laffer curve, since in this region, a small reduction in the tax rate could actually *raise* tax revenue, allowing the government to spend more on public goods anyway. Another way to think about this: for any level of tax revenue raised by a tax rate high enough to be on the wrong side of the tax rate, there is a *lower* tax rate that raises exactly the same tax revenue.

Since total tax revenue is initially increasing and finally decreasing in the tax rate, there is a tax rate that maximizes total revenue. Note that the tax rate that maximizes total tax revenue is not necessarily the tax rate that maximizes total welfare - that is determined by how much households value consumption of privately produced goods and services relative to goods and services produced by the government, and the extent of the distortionary effects of taxes. Empirically, substitution effects on labor supply are relatively small, so the tax rate at which the Laffer Curve peaks is relatively high.

6.2.5 Corporate Income Taxes

Corporate income taxes are usually imposed on corporate profits after variable expenses and depreciation are accounted for. How do corporate taxes affect investment, the capital stock and wages? We can answer this in partial equilibrium in our simple neoclassical framework.

With corporate income taxes, the profit maximization problem for a firm operating a Cobb-Douglas technology $Y = AK^{\alpha}N^{1-\alpha}$ is

$$\max_{K,N}(1-\tau_K)\left[AK^{\alpha}N^{1-\alpha} - wN - \delta K\right] - rK$$

The firm's first order condition with respect to capital gives

$$\underbrace{A\alpha\left(\frac{K}{N}\right)^{\alpha-1}}_{MPK} = \alpha\frac{Y}{K} = \underbrace{\frac{r}{1-\tau_K} + \delta}_{uc} \implies \frac{K}{Y} = \frac{\alpha}{\frac{r}{1-\tau_K} + \delta}$$

Clearly, in partial equilibrium, a higher tax rate will reduce the target K/Y ratio firms choose. This is because the higher tax rate reduces the profits firm owners can obtain by using a given amount of capital, and therefore makes earning the real interest rate by just saving the amount they were planning to invest in capital elsewhere. As firms invest less, the fall in the K/Y ratio raises the marginal product of capital until the point where the higher MPK offsets the higher capital tax rate and makes investment once more worthwhile.

Next, recall that in equilibrium, the real wage rate is the marginal product of labor, so that

$$w = (1-\alpha)A\left(\frac{K}{N}\right)^{\alpha} = (1-\alpha)A^{1/(1-\alpha)}\left(\frac{K}{Y}\right)^{\alpha/(1-\alpha)}$$

Since $0 < \alpha < 1$, the fall in K/Y must also lead to a fall in the real wage w. The lower level of capital stock leads to a lower marginal product of labor since capital and labor are complements in a Cobb-Douglas production function.

Finally, recall that investment I_t and the capital stock K_t are related by the relationship

$$\begin{aligned} K_{t+1} &= (1-\delta)K_t + I_t \\ \implies \frac{K_{t+1}}{Y_t} &= (1-\delta)\frac{K_t}{Y_t} + \frac{I_t}{Y_t} \\ \implies \frac{K_{t+1}}{Y_{t+1}}\frac{Y_{t+1}}{Y_t} &= (1-\delta)\frac{K_t}{Y_t} + \frac{I_t}{Y_t} \end{aligned}$$

Imposing steady state, where output grows at a constant rate g_Y, $\frac{K_{t+1}}{Y_{t+1}} = \frac{K_t}{Y_t} = \frac{K}{Y}$ and

$I_t/Y_t = I/Y$, we get

$$\frac{K}{Y}(1+g_Y) = (1-\delta)\frac{K}{Y} + \frac{I}{Y}$$
$$\implies \frac{I}{Y} = (g_Y + \delta)\frac{K}{Y}$$

Since K/Y falls, we must have that I/Y falls as well in steady state after an increase in corporate taxes.

6.3 Deficits and Public Debt Sustainability

6.3.1 Budget Deficits

Governments often finance their spending by issuing government debt, which is backed by the ability of the government to raise taxes in the future. The difference between current government spending and current tax revenue is sometimes called the budget deficit,

$$D_t = G_t - T_t$$

When a deficit is negative, the government is said to be running a budget surplus. Governments run deficits for a number of reasons:

- **Countercyclical fiscal policy**: Governments may choose to expand government spending in a recession and contract it during a boom to counter the effects of the business cycle. However, tax revenues are generally procyclical since the tax base, which includes all forms of taxable incomes, increase in booms and fall in recessions.
- **Tax Rate Smoothing**: If government spending rises for unanticipated reasons, it may choose to finance the sudden increases through debt. This minimizes how much taxes increase at one point in time, smoothing post-tax income.
- **Government Investments**: Government spending often pays off for society only in the long term, and hence funding these investments through taxes may be politically infeasible. Governments can finance investment projects through debt in these cases, which shifts the burden of paying for projects onto generations which benefit from them. This is the basis of the "1800's Rule", which states that the target deficit to (nominal) GDP level should be the ratio of government investment spending G^I to (nominal) GDP.

6.3.2 Long-Run Debt Levels

Let B_t be the outstanding stock of government debt in nominal terms, and let P_tY_t be nominal GDP. The deficit D_t represents the net addition to the debt, so that

$$B_{t+1} = B_t + D_t \implies D_t = \Delta B_t$$

Suppose nominal output grows at the constant rate g_{PY} and that in the long run, the ratio of deficits to GDP is a constant, i.e. $D_t/(P_tY_t) = D/PY$ for all t. Then note that

$$B_{t+1} = B_t + D_t$$
$$\implies \frac{B_{t+1}}{P_{t+1}Y_{t+1}}\frac{P_{t+1}Y_{t+1}}{P_tY_t} = \frac{B_t}{P_tY_t} + \frac{D_t}{P_tY_t}$$

Suppose we are in a steady state in the long run with a constant level of B/PY. Then, we must have

$$\frac{B}{PY}(1 + g_{PY}) = \frac{B}{PY} + \frac{D}{PY}$$
$$\implies \frac{B}{PY} = \frac{D}{PY}\frac{1}{g_{PY}}$$

6.3.3 Sustainable Debt and Public Debt Crises

When is a given long-run level of debt B/PY *sustainable*? For a given level of debt to be sustainable, the interest payments on it as a share of GDP, iB/PY, must at least be lower than tax revenues as a share of GDP, T/PY. Thus, we need

$$\frac{iB}{PY} \leq \frac{T}{PY} \implies \frac{B}{PY} \leq \frac{T/PY}{i}$$

However, governments tend to default on their debt - i.e. fail to meet interest payments on their outstanding debt - much before they hit this constraint. This occurs because in some cases, it can be *worth it* for governments to default. A defaulting government does not have to repay its debt today, and thus has more resources available for government spending and transfers. The tradeoff faced is usually exclusion from debt markets, which hampers the ability of the government to raise debt in the future and therefore to engage in stabilization policy, tax smoothing or government investment in the future.

Since governments can choose to default, bondholders are typically concerned about the possibility of being defaulted on. The concern that a government might be tempted to default, or to generate inflation that might reduce the value of nominal outstanding debt, is higher when the government is more indebted, i.e. it has a higher B/PY. However, when bondholders are more worried about default risk, they demand a *default premium* in order to hold government bonds, and the interest rates for government borrowing therefore rise. The higher the perceived risk of default, the higher the interest rate is.

Thus, concerns about debt sustainability can be self-fulfilling - if concerns about default lead to higher interest rates, this raises the cost of servicing interest rate repayments, which then makes default seem more tempting.

7 Economic Growth

In this section, we apply the neoclassical model we have developed in the last sections to study economy growth. We will start by discussing the Solow Model, which emphasizes the limits to economic growth imposed by diminishing returns, and then study the neoclassical model of economic growth. In the neoclassical model, growth in per-capita income will be driven entirely by growth in TFP, which we take as exogenous. The final part of this section will explore several models of endogenous growth, which try to understand the determinants of TFP growth itself. In the next section, we will study some of the facts surrounding economic growth.

7.1 Some Growth Facts

The following are a set of empirical regularities that govern the long-run behavior of macroeconomic aggregates.

- Over the long run, the share of labor and capital in GDP have remained roughly constant at about 65% and 35% respectively. Recently, some attention has been paid to a decline in the labor share.
- The capital to output ratio has been roughly constant.
- Capital and output per worker have grown at roughly constant rates.
- The capital stock itself has grown at a roughly constant rate.

7.2 The Solow-Swan Model

The Solow Model is one of the earliest general equilibrium models of economic growth. Prior to the Solow Model, it was believed that the key to economic growth was the accumulation of capital. This was based on the idea that economies were generally constrained by their available capital stock - i.e., they had an excess of labor - which meant that adding capital would allow the economy to grow at a steady rate, at least until capital was no longer scarce.

Solow's celebrated 1956 article argued that this logic was incomplete, due to diminishing returns to capital. He argued that for a given labor force, a higher capital stock would lower the rate of return to capital, thus cutting the growth in output achieved by raising capital. Eventually, raising capital indefinitely would raise output only by an infinitesimal amount, and economic growth would stop.

7.2.1 The Model

We now lay out Solow's model. We use lower-case letters for individual (per-capita) values and capital letters for aggregates. Consider an economy with N_t identical consumers. Each consumer inelastically supplies 1 unit of labor - that is, we assume that there is no disutility of working, and that consumers are endowed with one unit of labor. The population N_t (which equals the labor force) grows at the constant rate g_N, so that $N_t = N_0(1+g_N)^t$. Each consumer saves a constant fraction s of their income y_t and supplies one unit of labor. Thus, each consumer's consumption and saving are

$$c_t = (1-s)y_t \quad ; \quad s_t = sy_t$$

Adding over all consumers, aggregate consumption and savings are respectively

$$C_t = c_t N_t = (1-s)y_t N_t = (1-s)Y_t \quad ; \quad S_t = s_t N_t = sy_t N_t = sY_t$$

Consumers save by accumulating capital[46], so investment in this economy must equal savings: $I_t = S_t$. The total quantity of capital in the economy evolves according to

$$K_{t+1} = (1-\delta)K_t + I_t = (1-\delta)K_t + S_t = (1-\delta)K_t + sY_t$$

A representative firm uses capital K_t and the N_t units of labor available to produce output $Y_t = K_t^\alpha (Z_t N_t)^{1-\alpha}$. Note that technology Z_t is "Labor Augmenting": growth in Z allows a given number of workers to provide a larger amount of effective labor services, and therefore represents an increase in the number of "efficiency" units of labor. In this economy, TFP A_t is just given by[47] $A_t = Y_t/(K_t^\alpha N_t^{1-\alpha}) = Z_t^{1-\alpha}$.

The Solow Model is an **exogenous** growth model, so we will assume that Z_t grows at a constant rate g_Z, i.e. that $Z_{t+1} = Z_t(1+g_Z)$. You should convince yourself that in the Cobb-Douglas case, this is isomorphic to assuming a constant growth rate of TFP A_t - specifically, to assuming that A_t grows at the rate $g_A = (1-\alpha)g_Z$.

Given initial levels Z_0, N_0, K_0 of labor productivity, labor and capital respectively and exogenous growth rates of population g_N and technology g_Z (equivalently, an exogenous growth rate for TFP $g_A = (1-\alpha)g_Z$), an **Equilibrium** in the Solow Model is a set of paths for output, consumption, capital and investment Y_t, C_t, K_t, I_t satisfying the following:

- Consumption is a constant fraction $1-s$ of output, so that $C_t = (1-s)Y_t$.
- Capital evolves according to the equation $K_{t+1} = (1-\delta)K_t + I_t$.

[46] Technically, the assumption here is that physical capital is the only asset in positive net supply, as we discussed in section 4. For example, if we introduce a banking sector which collects "deposits" from households and lends these out to firms to fund capital accumulation, such that total loans equals total deposits, then the equation $S_t = I_t$ still holds since deposits and loans are equal and offset each other.

[47] The fact that we can represent labor-augmenting technical change, i.e. growth in Z, as isomorphic to TFP growth is a feature of the Cobb-Douglas production function, and is not true in general for more complicated production functions.

- The goods market clears, i.e. $I_t = Y_t - C_t = sY_t$.
- Output and inputs are related via the production function

$$Y_t = K_t^\alpha (Z_t N_t)^{1-\alpha}$$

7.2.2 Digression: Endogenous vs Exogenous Variables

Recall that a model can be considered as a system of equations which, given values of *exogenous* variables, can be solved for *endogenous* variables. Let's interpret quantities in the Solow Model through this lens.

7.2.3 Equilibrium

Let's characterize the behavior of the economy in an equilibrium. Start with the equation for the evolution of the capital stock. We have,

$$K_{t+1} = (1-\delta)K_t + sY_t = (1-\delta)K_t + sK_t^\alpha (Z_t N_t)^{1-\alpha}$$

To solve for the path for capital, it is convenient to normalize all variables by the number of efficiency units of labor input $Z_t N_t$. Define $k_t = K_t/(Z_t N_t)$. We have,

$$\begin{aligned}
&\frac{K_{t+1}}{Z_{t+1}N_{t+1}} \frac{Z_{t+1}N_{t+1}}{Z_t N_t} = (1-\delta)\frac{K_t}{Z_t N_t} + s\frac{K_t^\alpha (Z_t N_t)^{1-\alpha}}{Z_t N_t} \\
\implies\ & k_{t+1}(1+g_N)(1+g_Z) = (1-\delta)k_t + sk_t^\alpha
\end{aligned}$$

We define a **Balanced Growth Path** (BGP) as an Equilibrium of a model in which output, capital, and consumption grow at a constant and equal rate, so that the ratio of consumption to output (in the Solow Model, this is just $1-s$) and the ratio of capital to output are constant over time. Note that the latter ratio is just

$$\frac{K_t}{Y_t} = \frac{K_t}{K_t^\alpha (Z_t N_t)^{1-\alpha}} = k_t^{1-\alpha}$$

and hence, along a BGP, we must have $k_t = k_{t+1} \equiv k^*$. We can characterize a BGP of the Solow Model by finding the steady-state value of k^* that is consistent with the dynamic equation for capital, and then solve for the remaining quantities in the economy as a function of k^*. We have,

$$\begin{aligned}
& k^*(1+g_N)(1+g_Z) = (1-\delta)k^* + s{k^*}^\alpha \\
\implies\ & k^*(g_N + g_Z + g_N g_Z + \delta) = s{k^*}^\alpha \\
& \implies k^* = \left(\frac{s}{g_N + g_Z + \delta + g_N g_Z}\right)^{\frac{1}{1-\alpha}}
\end{aligned}$$

For small growth rates of TFP and population, the term $g_N g_Z$ is negligible, and we can just write

$$k^* = \left(\frac{s}{g_N + g_Z + \delta} \right)^{\frac{1}{1-\alpha}}$$

This equation shows that the steady state level of capital per efficiency units of labor is

- increasing in the saving rate, since more saving leads to more capital accumulation for a given depreciation rate
- decreasing in the depreciation rate, as a higher depreciation rate leads to lower useable capital surviving from period-to-period for a given saving rate. A higher depreciation rate implies a lower level of *net* investment for a given level of *gross* investment, the latter of which equals savings.
- the growth rate of efficiency units of labor, $g_N + g_Z$. The faster efficiency units of labor grow, the more thinly spread a given amount of new capital is across each unit of labor.

Given k^*, we can easily calculate output per efficiency unit of labor and consumption per efficiency unit of labor, using

$$\begin{aligned} y^* &= k^{*\alpha} \\ c^* &= (1-s)k^{*\alpha} \end{aligned}$$

Consumption per capita, which is what matters for welfare, is just

$$c_t = \frac{C_t}{N_t} = c^* Z_t = Z_t (1-s) k^{*\alpha}$$

which clearly grows at the rate g_Z, since c^* is constant along the BGP. Somewhat surprisingly, in the long run, all growth in consumption per capita comes from technological progress, i.e. g_Z! This is the key insight of the Solow Model: in the absence of technical progress, with a constant saving rate, growth must come to a stop.

Why is this the case? The key is diminishing returns to capital: as firms accumulate more and more capital per efficiency unit of labor, the returns to these investments on the margin fall. This continues until eventually, adding enough capital to raise the capital-efficiency labor ratio by an extra unit generates just enough extra output per efficiency labor units to cover the depreciation of that unit between periods. At this point, the capital to efficiency unit of labor ratio stabilizes.

This is most easily visible in the Cobb-Douglas case. We know that under Cobb-Douglas,

$$\begin{aligned} Y_t &= K_t^{\alpha} (Z_t N_t)^{1-\alpha} \\ \implies \frac{Y_{t+1}}{Y_t} &= \left(\frac{K_{t+1}}{K_t} \right)^{\alpha} \left(\frac{Z_{t+1} N_{t+1}}{Z_t N_t} \right)^{1-\alpha} \\ \implies \log(1+g_Y) &= \alpha \log(1+g_K) + (1-\alpha)\log(1+g_N) + (1-\alpha)\log(1+g_Z) \\ \implies g_Y &= \alpha g_K + (1-\alpha)(g_N + g_Z) \end{aligned}$$

where the final line follows from the approximation $\log(1+x) \approx x$. Subtract g_N from both sides to get

$$g_{Y/N} = (1-\alpha)\, g_Z + \alpha\, g_{K/N} \equiv g_A + \alpha g_{K/N} \tag{12}$$

Thus, growth in output per worker (and hence in consumption per worker along a balanced growth path) is driven by growth in labor productivity and growth in capital per worker (which is sometimes called "capital deepening."). We can go further by expressing the K/N ratio in terms of K/Y. Under Cobb Douglas production, we have

$$\begin{aligned}
Y_t &= K_t^{\alpha}(Z_t N_t)^{1-\alpha} \\
\implies \frac{K_t}{Y_t} &= \left(\frac{K_t}{Z_t N_t}\right)^{1-\alpha} \\
\implies \frac{K_t}{N_t} &= Z_t \left(\frac{K_t}{Y_t}\right)^{\frac{1}{1-\alpha}} \\
\implies g_{K/N} &= g_Z + \frac{g_{K/Y}}{1-\alpha}
\end{aligned}$$

Substituting this into equation 12,

$$g_{Y/N} = g_Z + \frac{\alpha}{1-\alpha}\, g_{K/Y} \equiv \frac{g_A}{1-\alpha} + \frac{\alpha}{1-\alpha}\, g_{K/Y}$$

From the capital accumulation equation, we know that

$$\begin{aligned}
K_{t+1} &= (1-\delta)K_t + sY_t \\
\implies \frac{K_{t+1}}{Y_{t+1}}(1+g_Y) &= (1-\delta)\frac{K_t}{Y_t} + s \\
\implies \frac{K_{t+1}/Y_{t+1}}{K_t/Y_t} &= \frac{1-\delta}{1+g_Y} + \frac{s}{(1+g_Y)K_t/Y_t} \\
\implies 1 + g_{K/Y} &= \frac{1-\delta}{1+g_Y} + \frac{s}{(1+g_Y)K_t/Y_t} \\
\implies g_{K/Y} &= \frac{s}{(1+g_Y)K_t/Y_t} - \frac{g_Y+\delta}{1+g_Y}
\end{aligned}$$

Consider a balanced growth path, along which g_K, g_Y are constant (and so $g_{K/Y}$ is constant as well.). Suppose $g_{K/Y}$ was positive at each date in an equilibrium. Then K/Y would be increasing over time, which would drive the first term on the right side to zero eventually. The right side would then have a negative value, which is a contradiction to the assumption that the left side is positive! Thus, in the long run, it must be that $g_{K/Y} = 0$. Equation 12 then shows that the only source of long run balanced growth must be technical progress!

7.2.4 The Golden Rule

Note that the simple Solow Model does not involve any optimizing decisions by any agent in the economy. In particular, the saving rate in the model is taken as exogenous.

We now ask: for what saving rate will consumers be best off in the Solow Model? Since consumers do not suffer disutility from labor (by the inelastic labor supply assumption), this corresponds to the choice of the saving rate that maximizes consumption per capita. Since consumption per capita is proportional to c^*, we just need to find the saving rate that maximizes c^*. We have,

$$c^* = (1-s)k^{*\alpha} = (1-s)\left(\frac{s}{g_N + g_Z + \delta}\right)^{\frac{\alpha}{1-\alpha}}$$

We could just calculate first-order conditions and set $dc^*/ds = 0$, but it's particularly convenient to do one step in between - take logs. We have,

$$\begin{aligned} & \log c^* = \log(1-s) + \frac{\alpha}{1-\alpha}\log s - \frac{\alpha}{1-\alpha}\log(g_N + g_Z + \delta) = 0 \\ \implies & \frac{1}{c^*}\frac{dc^*}{ds} = -\frac{1}{1-s} + \frac{\alpha}{1-\alpha}\frac{1}{s} = 0 \\ \implies & s = \alpha \end{aligned}$$

so the optimal saving rate is equal to the capital share parameter of the economy. To gain some more insight, consider the value of k corresponding to this saving rate (call it k_g). Noting that

$$c^* = \underbrace{f(k^*)}_{\text{output per eff. lab. unit}} - \underbrace{(\delta + g_N + g_Z)k^*}_{\text{investment per eff. lab. unit}}$$

, the value of capital per efficiency labor unit that maximizes consumption must solve

$$f'(k_g) = \delta + g_N + g_Z$$

The left side of this equation is the increase in output at the margin for each unit of capital invested. The right side is the required increase in investment that is necessary to maintain the level of capital k_g per efficiency labor unit. If the left side is larger, it is possible to increase consumption by raising k, since the investment required to maintain the increased k is smaller than the marginal rise in output.

7.2.5 Applying the Solow Model

The Solow Model can be used to study the impact of shocks to the economy on the growth rates and levels of output, consumption and investment. In order to apply the Solow model to study changes, a general recipe is to proceed as follows.

- The Solow model takes as given the growth rates of productivity and labor, g_Z and g_N, and the saving rate $s = I/Y$. Note that if the growth rate of TFP g_A is given, then g_Z can be easily calculated using $g_Z = g_A/(1-\alpha)$.
- Use the equation

$$g_Y = g_N + g_{Y/N} = g_N + g_Z = g_N + \frac{g_A}{1-\alpha}$$

  to determine the growth rate of total output. Along a BGP, this growth rate is also the growth rate of C, I, K.

- Along a BGP, we know that

$$\frac{K_t}{Y_t} = \frac{I_t/Y_t}{g_Y + \delta} = \frac{s}{g_Y + \delta}$$

  Use this equation to determine the K/Y ratio.

- Given the K/Y ratio it is straightforward to calculate the Y/N ratio using the production function.

- Once all the ratios have been determined, noting that N is exogenous, determine the impacts on the levels of Y, I, K. The impact on the level of consumption can be backed out using $C = Y - I$.

Let's apply the Solow model to study the impact of a demographic transition that raises the population growth rate permanently.

- **First**, note that the increase in g_N will, in the long run, raise g_Y - but not change $g_{Y/N}$, since the latter only depends on g_Z.

- **Second**, there is no change in s, so from the equation

$$\frac{K_t}{Y_t} = \frac{s}{g_Y + \delta}$$

  we know that the K/Y ratio must fall in the long run.

- **Third**, note that K/Y and Y/N are positively related (show this!), so in the long run the decline in K/Y permanently shifts the path of Y/N lower. However, on impact, there is no decline in the level of any variable - it's just that the rate of capital accumulation per efficiency labor unit slows down temporarily. In the long run, the growth rate of Y/N goes back to g_Z.

- **Finally**, since $C = (1-s)Y$, we have $C/N = (1-s)Y/N$, which shows that consumption per capita falls as well relative to the old BGP. It is not the case that C/N falls in absolute terms - there is a temporary slowdown in the growth of C/N, but eventually growth in C/N goes back to its original level, g_Z.

7.3 The Neoclassical Growth Model

The Solow Model treats the rate of capital accumulation, equivalently the rate of saving, as exogenous. We now endogenize the rate of saving - that is, we allow households to choose the rate at which they save. The model we will obtain by doing this is called the neoclassical growth model, whose core ideas were described as early as Ramsey (1928).

7.3.1 The Model

The neoclassical growth model's structure is similar to that of the Solow Model. Once again, consider an economy with N_t identical consumers, each of whom inelastically supplies 1 unit of labor. The labor force thus grows at the constant rate g_N, so that $N_t = N_0(1+g_N)^t$. Consumers begin life with initial capital stock K_0, and consume an amount c_t per capita each period[48]. Their choice of consumption solves an infinite-horizon intertemporal consumption-saving problem,

$$\max_{\{c_t\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \frac{1}{(1+\rho)^t} u(c_t)$$
$$\text{subject to}$$
$$c_t + K_{t+1} = (1+r_t)K_t + w_t N_t$$

where w_t, r_t are the real wage rate and real interest rate at date t, and $u(\cdot)$ is an increasing and strictly concave felicity function. We will work with the constant elasticity of substitution specification, $u(c) = c^{1-1/\sigma}/(1-1/\sigma)$. Note that we have assumed that households save by accumulating capital.

We have seen this problem before - it is analogous to the models studied in section 4.2.2, except for the infinite horizon. The solution to this problem is still characterized by the Euler Equation

$$u'(c_t) = \frac{1+r_{t+1}}{1+\rho} u'(c_{t+1})$$

and a technical condition required to rule out explosive solutions[49]. Given isoelastic utility, we have

$$c_t^{-1/\sigma} = \frac{1+r_{t+1}}{1+\rho} c_{t+1}^{-1/\sigma} \tag{13}$$

We next consider firms. Firms rent capital at rental rate r_t from households and hire labor at a wage rate of w_t. They solve the static problem

$$\max_{K_t, N_t} K_t^{\alpha}(Z_t N_t)^{1-\alpha} - \delta K_t - w_t N_t - r_t K_t$$

where δ is the depreciation rate and Z_t is the level of labor productivity, which we assume grows at the constant rate g_Z (recall that TFP in this economy is $A_t = Z_t^{1-\alpha}$, so a constant growth rate g_Z of Z_t just corresponds to a constant growth rate $g_A = (1-\alpha)g_Z$ for TFP A_t. The solution to this problem is characterized by the two first-order conditions

$$MPK_t \equiv \alpha \frac{Y_t}{K_t} = \alpha \left(\frac{K_t}{Z_t N_t}\right)^{\alpha-1} = r_t + \delta \tag{14}$$

$$MPN_t \equiv (1-\alpha)\frac{Y_t}{N_t} = (1-\alpha) Z_t \left(\frac{K_t}{Z_t N_t}\right)^{\alpha} = w_t \tag{15}$$

48 That is, each consumer consumes c_t, so total consumption is $C_t = N_t c_t$.

49 Technically, we need to impose a "No-Ponzi" condition - that is, we require that the present value of household net worth is weakly positive at dates far enough into the future.

Finally, we consider market clearing. The model has three markets: those for capital, labor and goods. Capital and labor market clearing just impose that the quantity of capital demanded by firms equals the amount of capital households entered the period with, and that the amount of labor demanded by firms equals the inelastic supply of labor N_t. Goods market clearing just requires that total output produced, $Y_t = A_t K_t^{\alpha} N_t^{1-\alpha}$, equals total consumption plus total gross investment, so that

$$Y_t = C_t + I_t = C_t + K_{t+1} - (1-\delta)K_t \tag{16}$$

Given initial values Z_0, N_0, K_0 for productivity, labor and capital and given the growth rates of productivity and labor supply g_Z, g_N, an **Equilibrium** of the neoclassical growth model is a set of paths $\{C_t, K_t, Y_t\}_{t=0}^{\infty}$ for real quantities consumption, capital and output and real prices w_t, r_t that satisfy:

- Optimal household choices for consumption and saving, i.e. the Euler Equation 13[50].
- Optimal firm choices for capital and labor, i.e. the capital and labor FOCs 14 and 15.
- Market clearing in goods, labor and capital markets.

7.3.2 Characterizing a Balanced Growth Path

A **Balanced Growth Path** (BGP) is an Equilibrium of the neoclassical growth model along which

- Z, N, Y all grow at a constant rate, and
- the ratios $C/Y, K/Y$ and I/Y are constant.

Let's characterize a BGP. In doing this, we will try to express as many of the endogenous quantities in our model as functions of exogenous quantities and parameters only.

Start with the easiest of the quantities: labor. We know by labor market clearing that in equilibrium, the amount of labor hired must equal exogenously given labor supply N_t. Thus, we don't have to do anything more: given an initial amount of labor and an exogenous growth rate g_N for labor, we know the entire path of labor input at all dates already.

Next, as in the Solow Model, as long as the technology is Cobb-Douglas, we must have

$$g_{Y/N} = g_Z + \frac{\alpha}{1-\alpha} g_{K/Y}$$

[50]... And a condition to rule out over-borrowing.

Along a BGP, we know that K, Y must grow at the same rate, which means that the ratio K/Y must be constant. Thus, $g_{K/Y} = 0$, and we have

$$g_{Y/N} = g_Z = \frac{g_A}{1-\alpha} \tag{17}$$

which establishes the growth rate of output per capita as a function of exogenous variables and parameters only.

First, we will use the household's optimality conditions to establish a value for the user cost of capital. Second, given the user cost of capital, we will use the firm's first order conditions to back out what the optimal K/Y ratio must be. Third, we will use the law of motion for capital and the goods market and labor market clearing conditions to establish the level of output and consumption consistent with this.

Start with the household. The Euler equation 13 can be rearrange to give us

$$\left(\frac{c_{t+1}}{c_t}\right)^{1/\sigma} = \frac{1+r_{t+1}}{1+\rho} \implies 1 + g_{ct} = \left(\frac{1+r_{t+1}}{1+\rho}\right)^{\sigma}$$

where g_{ct} is the growth rate of per-capita consumption. Along a BGP, we know that g_c is a constant, so we drop the t subscript. Taking logs and applying the approximation $\log(1+x) \approx x$, we have

$$g_c = \sigma(r_{t+1} - \rho) \implies r_{t+1} = \rho + \frac{1}{\sigma} g_c$$

which shows that the real interest rate r is constant on a BGP.

By definition, on a BGP, aggregate consumption and output grow at the same rate, so per-capita output and consumption must also grow at the same rate. Thus, $g_c = g_{Y/N}$. By equation 17, we thus have $g_c = g_Z = g_A/(1-\alpha)$. Thus, we know that

$$r = \rho + \frac{g_Z}{\sigma} = \rho + \frac{1}{\sigma}\frac{g_A}{1-\alpha} \tag{18}$$

which expresses the real interest rate as a function of exogenous variables and parameters only.

Next, let's turn to the capital stock. In the version of the model we have described here, the user cost of capital is just $r + \delta$, since we have abstracted from capital taxes or changes in the relative price of capital. The firm's first order condition for capital 14 gives us

$$r + \delta = \alpha \frac{Y_t}{K_t} \implies \frac{K_t}{Y_t} = \frac{\alpha}{r+\delta}$$

which establishes that the K/Y ratio is indeed constant, as required by the BGP. Substituting for r from 18, we have

$$\frac{K_t}{Y_t} = \frac{\alpha}{\rho + \delta + \frac{g_Z}{\sigma}}$$

Under Cobb-Douglas, we know that

$$\frac{K_t}{Y_t} = \left(\frac{K_t}{Z_t N_t}\right)^{1-\alpha}$$

so we have,

$$\frac{K_t}{Z_t N_t} = \left(\frac{\alpha}{\rho + \delta + \frac{g_Z}{\sigma}}\right)^{\frac{1}{1-\alpha}}$$

Recall that the growth rates g_Z, g_N are exogenously given, so given initial values Z_0, N_0 the paths Z_t, N_t are fully determined. Thus, this equation can be used to determine the value of K_t at any date as a function of exogenous variables and parameters only. Given K_t, N_t, Z_t, we can also pin down output Y_t using the production function and investment I_t using the capital transition equation $K_{t+1} = (1-\delta)K_t + I_t$. We have,

$$\begin{aligned} I_t &= K_{t+1} - (1-\delta)K_t \\ \implies \frac{I_t}{Y_t} &= \frac{K_{t+1}}{Y_{t+1}}(1+g_Y) - (1-\delta)\frac{K_t}{Y_t} \\ &= (g_Y + \delta)\frac{K_t}{Y_t} \\ &= (g_N + g_Z + \delta)\,\frac{\alpha}{\rho + \delta + \frac{g_Z}{\sigma}} \end{aligned}$$

where the third line uses the fact that on the BGP $K_{t+1}/Y_{t+1} = K_t/Y_t$. This expresses the investment-output ratio as a function of exogenous variables and parameters only.

Finally, once investment has been pinned down, consumption can be backed out from the goods market clearing condition, $C_t = Y_t - I_t$.

7.3.3 Applying the Neoclassical Growth Model

Just like the Solow Model, we can use the Neoclassical Growth Model to study the impacts of shocks on the growth rates and levels of macroeconomic aggregates in the long run. Using the neoclassical growth model emphasizes the importance of endogenizing the saving rate in determining the impact of a shock to the economy on the growth rate of output per worker and on its levels. In order to apply the neoclassical growth model, a general recipe is to proceed as follows.

- The neoclassical growth model takes as given the growth rates of productivity and labor, g_Z and g_N. Note that if the growth rate of TFP g_A is given, then g_Z can be easily calculated using $g_Z = g_A/(1-\alpha)$.

- Use the equation

$$g_Y = g_N + g_{Y/N} = g_N + g_Z = g_N + \frac{g_A}{1-\alpha}$$

to determine the growth rate of total output and output per capita. Along a BGP, this latter growth rate is also the growth rate of $C/N, I/N, K/N$.

- Use these growth rates to calculate the real interest rate, using the Euler equation, and determine the impact on the real interest rate of any changes in the economy. Infer the impact on the user cost of capital.

- Use the firm's first order condition for capital to obtain the impact on the K/Y ratio.

- Given the K/Y ratio it is straightforward to calculate the I/Y ratio and the Y/N ratio using the law of motion for capital and the production function.

- Once all the ratios have been determined, noting that N is exogenous, determine the impacts on the levels of Y, I, K. The impact on the level of consumption can be backed out using $C = Y - I$.

As we did with the Solow Model, let's apply the model to study the impact of a demographic transition that raises the population growth rate permanently.

- **First**, note that (exactly as in the Solow Model), the increase in g_N will, in the long run, raise g_Y - but not change $g_{Y/N}$, since the latter only depends on g_Z.

- **Second**, note that with no change in $g_{Y/N}$, the real interest rate is unchanged. Since the depreciation rate doesn't change either, there is no long-run change in the user cost of capital.

- **Third**, with no change in the user cost of capital, we know that there is no change in the K/Y ratio. From the equation

$$\frac{I_t}{Y_t} = (g_Y + \delta)\frac{K_t}{Y_t}$$

we know that the I/Y ratio must therefore have gone up.

- **Fourth**, since there is no long-run change in K/Y, there is no impact on the long-run growth path of Y/N, which continues to grow at the rate g_Z.

- **Finally**, since C/N and Y/N both grow at the same rate, there is no impact on the growth rates of consumption per capita.

Note that this is a *very* different conclusion from the Solow Model's analysis, and the key to understanding this difference is to note that the saving rate is endogenous in the Neoclassical Growth Model. In the Solow Model, the savings rate does not respond to the growth in labor force, so the rise in g_N implies a more rapid growth in labor than in capital, leading to a fall in the K/N ratio and, in turn, in the K/Y ratio. In the Neoclassical growth model, agents respond to rapid population growth by raising their savings rate correspondingly, maintaining the same K/Y ratio.

7.4 Endogenous Growth Models based on Human Capital

In the Solow Model and the Neoclassical Growth Model, the growth rate of per-capita quantities in the economy is determined by the growth of labor productivity Z_t (or equivalently, in the Cobb-Douglas case, by growth in TFP A_t). But the growth rate of productivity g_Z is taken as an exogenous quantity. Thus, while these models establish that sustained growth in per-capita output is only possible via productivity growth, they are not a complete theory of growth since they do not pin down where TFP growth comes from.

Understanding TFP growth is the subject of *Endogenous* growth models. These models resemble the neoclassical growth model in many ways - in particular, they involve households choosing how much to save and consume optimally and firms choosing investment optimally to hit a target capital-output ratio - but they add a set of equations that can be used to calculate the growth rate of productivity.

7.4.1 The Lucas Model

The first endogenous growth model we study emphasizes the role of human capital. In this interpretation, rising human capital per worker is the driving force behind growth in labor productivity. Mankiw (1991) CITE argues that the appropriate way to think of human capital is as the stock of knowledge that has been transmitted from the sum total of technological and scientific discoveries made up to each point in time to the brains of human workers via education.

Accordingly, we develop the Lucas model as follows. As in the Lucas and Solow Models, we start with consider an economy with N_t identical consumers, each of whom inelastically supplies 1 unit of labor. We assume that each period, each consumer spends a fraction s_E of its time accumulating human capital (this should be thought of as the fraction of time individuals spend in school or in on-the-job training).

Education allows individuals to accumulate human capital h_t per worker. In particular, individuals accumulate human capital according to the law of motion

$$h_{t+1} = h_t + Z_E s_E h_t \tag{19}$$

where Z_E should be interpreted as the efficiency with which time and human capital devoted to school is converted to human capital. Note that we can write

$$g_h = \frac{h_{t+1} - h_t}{h_t} = Z_E s_E$$

which is constant over time; this will be crucial for the existence of a Balanced Growth Path.

How does human capital translate into production? We model this by assuming that a consumer with human capital h_t can supply $(1 - s_E)h_t$ efficiency units of labor. Thus, human capital works by increasing the effective labor input for each hour worked.

As in the neoclassical model, consumers begin life with initial capital stock K_0 and initial human capital h_0, and consume an amount c_t per capita each period. Their choice of consumption solves the same infinite-horizon intertemporal consumption-saving problem, except for their labor income.

$$\max_{\{c_t\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \frac{1}{(1+\rho)^t} u(c_t)$$
$$\text{subject to}$$
$$c_t + K_{t+1} = (1+r_t)K_t + w_t(1-s_E)h_t N_t$$

where w_t, r_t are the real wage rate and real interest rate at date t, and $u(\cdot)$ is the constant elasticity of substitution felicity function $u(c) = c^{1-1/\sigma}/(1-1/\sigma)$. The solution to this problem is characterized by the same Euler equation as above, and a similar set of conditions to rule out explosive solutions.

Let $\tilde{N}_t$ be the total input of efficiency units of labor. Firms hire efficiency units of labor $\tilde{N}_t$ and solve the problem

$$\max_{K_t, \tilde{N}_t} AK_t^{\alpha}\tilde{N}_t^{1-\alpha} - w_t\tilde{N}_t - r_t K_t$$

which, as usual, has the solution characterized by the firm's first order conditions,

$$A\alpha \left(\frac{K_t}{\tilde{N}_t}\right)^{\alpha-1} = r_t$$
$$A(1-\alpha)\left(\frac{K_t}{\tilde{N}_t}\right)^{\alpha} = w_t$$

Finally, market clearing in the capital market requires that capital supply equal capital demand, as in the neoclassical model, and market clearing in the goods market requires that $Y_t = C_t + K_{t+1} - (1-\delta)K_t$. Note that while investment in physical capital shows up in the goods market clearing condition, investment in human capital does not. This is a result of the assumption that human capital accumulation only requires time, and not actual physical investment (say, in constructing educational institutions or buying supplies for students).

In the labor market, the total demand for efficiency units of labor $\tilde{N}_t$ must equal the total supply of efficiency units of labor. Since each consumer works for $(1-s_E)$ hours and there are N_t consumers, the total number of hours worked is $(1-s_E)N_t$. Each hour worked is worth h_t efficiency units of labor. Thus, for labor market clearing, we must have $\tilde{N}_t = (1-s_E)h_t N_t$.

Given a path for the population $\{N_t\}$ and values for $K_0, h_0, s_E, Z_E)$, an **Equilibrium** of the model is a set of paths for consumption, capital, output and labor satisfying the household's optimality conditions, the firm's optimality conditions, the laws of motion for physical and human capital 19 and the market clearing conditions for labor, capital and output. A **Balanced Growth Path** is an equilibrium along which consumption, investment, output and the capital stock all grow at the same rate.

7.4.2 Characterizing a Balanced Growth Path

We can characterize the growth rate of per-capita output in the economy along a BGP by combining the law of motion for human capital, the labor market clearing condition and the production function. For simplicity, we assume that the population is constant[51], that is, $N_t = N$ for all dates t. Recall that

$$g_h = \frac{h_{t+1} - h_t}{h_t} = Z_E s_E$$

From the production function, it can be shown that[52]

$$\frac{Y_t}{N} = A^{1/(1-\alpha)} \left(\frac{K_t}{Y_t}\right)^{\alpha/(1-\alpha)} (1 - s_E) h_t$$

Taking logs,

$$\log(Y_t/N) = \log\left[A^{1/(1-\alpha)} \left(\frac{K_t}{Y_t}\right)^{\alpha/(1-\alpha)} (1 - s_E)\right] + \log h_t$$
$$\implies g_{Y/N} = g_h = Z_E s_E$$

where the final line follows from the fact that along a BGP, K/Y is a constant. The economy thus grows, even without exogenous TFP growth! All growth instead comes from the *endogenous* accumulation of human capital. Observe that changes in the efficiency of human capital accumulation Z_E have *growth* effects on the economy, since they change the growth rate of the economy on impact, but no *level* effects, since they do not affect the level of output on impact. By contrast, a change in s_E has a growth and a level effect, since an increase in s_E today will raise the growth rate of the economy at the cost of output today due to the lower level of labor input on impact.

Note that the structure of the model is similar to the structure of the neoclassical growth model, so that once the growth rate of output per worker is pinned down, we can just follow the steps outlined in section 7.2.3 to pin down the behavior of the level of output, capital, investment and other macroeconomic aggregates in response to shocks to the economy.

7.4.3 Challenges to Human Capital based theories of Growth

One implication of theories of growth emphasizing human capital accumulation is that the higher the level of resources (in the simple model, time) devoted to human capital acquisition, the higher the growth rate of the economy (and not just the level of output in the economy). However, globally, it turns out that average years of schooling has been trending up, while the growth rate of output per worker has not increased.

[51] The extension to the case with positive population growth is an exercise for the reader!
[52] Show this.

Another way to see this is that across countries, higher levels of average education, a common proxy for human capital, are correlated with higher levels of income per capita but not with the growth rate of income per capita. Thus, in the long run, it seems that much like investment in physical capital, a higher human capital-output ratio leads to higher output per worker levels, but not higher growth rates - that is, human capital runs into diminishing returns as well.

7.5 Endogenous Growth Models based on the Creation of Ideas

The "idea" that economic growth is the result of the creation of ideas rather than capital accumulation was popularized in a famous paper by Paul Romer (1990). In this paper, Romer argues that the set of economic goods should be partitioned into a set of objects and a set of *ideas*, which consist of instructions and recipes for performing economic activities, including production, distribution and even consumption. The set of ideas includes not just engineering plans and blueprints, but also organizational innovations like just-in-time inventory management and basic science (including the approximation $\log(1+x) \approx x$ we've grown to love!).

7.5.1 Nonrivalry and Increasing Returns

Modern economic growth theory starts with two observations about ideas. First, the set of all possible ideas is, for all practical intents and purposes, unlimited - far more unlimited than the material constraints on resources on a finite planet. Thus, one way to sustain growth with finite resources is the continued creation of new ways to combine finite resources.

Second, most objects are *rivalrous in consumption* - the consumption of a good by one individual precludes the consumption of that good by any other individual. As a result, most goods and services studied in Economics are *scarce* - they are in finite supply, relative to unlimited wants[53]. However, ideas are fundamentally different from most goods: they are *nonrival in consumption*. Once an idea has been created, it can be taught, understood and applied by any number of individuals without any impact on the ability of each of those individuals to use the idea. That is, one individual's "consumption" of an idea does not reduce the "amount" of the idea available to others.

To fix "ideas", let's consider some examples of ideas: differentiation in calculus, the guitar tabs for *Stairway to Heaven* and the blueprint for producing a MacBook.

- It should be clear that the *idea* of how to take a derivative is nonrival - odds are that at any moment of time, an enormous number of people are taking a derivative, but the exact number is irrelevant to quantifying the benefit of taking

[53]Note that throughout this course, the isoelastic utility function we use has a positive marginal product for any positive level of consumption. This is the notion of "unlimited wants" - no matter what the amount of consumption, the positive marginal product means that the consumer will always prefer higher consumption to less.

that derivative for any individual. But the textbooks that contain the details of *how* to take a derivative are rival in consumption - it is impossible for two people to share a given calculus textbook and derive the same utility that they would have if they each had their own copy of the book[54].

- Any number of Led Zeppelin cover bands can probably play *Stairway to Heaven* simultaneously - given that the song has been composed, no single band's performance affects the ability of any other band to play the song. While guitars, drums and performance venues (and talented players!) are all scarce, the idea embodied in the composition itself is unaffected when played.
- Once the blueprint for a new MacBook is complete, Apple doesn't need to invent a new blueprint for every MacBook it produces - rather, each computer now embodies the design that was created originally. While it may be necessary to duplicate the digital files or physical sets of documents that codify the blueprint (paper and computer space are rival in consumption), the production of copies of the idea does not require the idea to be re-invented.

At this point, it's worth pausing and asking - well, if the blueprints to a MacBook are nonrival in consumption, why doesn't *everyone* produce MacBooks? This leads to the concept of *excludability*, which measures the extent to which individuals can assert property rights over goods and ideas. Note that nonrivalry and excludability are distinct concepts. Nonrivalry is a statement about the feasibility of multiple people using a good or idea simultaneously, while Excludability measures the extent to which it is possible to establish ownership over a good or idea and *exclude* others from using it. Some ideas are nonexcludable - it is impossible to prevent people from using differentiation in their research or at work - while others are excludable, via the patent system or other intellectual property rights legislation. The latter category includes most production processes and new products protected by trademarks and copyrights.

Why is nonrivalry so important? This is because it implies that income per person depends on the aggregate stock of ideas, not the stock of ideas per person. This is in contrast with capital - output per person depends on capital per worker, as we've seen in the models above! Unlike machines or factories, a new idea can be used by any number of people at once, so every improvement has the potential to benefit everyone. Raising individual productivity through capital accumulation requires an increase in capital *per worker*, but once an idea for how to use a computer is invented (say, someone invents Python), this idea can be used by anyone with access to a computer simultaneously.

More formally, nonrivalry can lead to increasing returns to scale. To see this, let's start with the standard argument that justifies constant returns to scale, the **replication argument**, which goes as follows. Suppose we know that a factory that uses input l, k for labor and capital produces output y. Consider constructing an identical factory next to the old one and allocating an identical set of inputs k, l to each factory. Then, the total

[54] In a world with eBooks, it may seem like this isn't quite true when you can always just share a copy. Note that the book is still rival though: two copies of the textbook take up twice as much computer memory as a single copy does, and therefore in order for two individuals to enjoy an individual copy still requires twice as many resources as a single copy would.

output of the combination must be $2y$, and the total inputs are just $2k$ and $2l$. Thus, the relationship between inputs and output must satisfy constant returns. An important implication of this is that the cost of production of output y is just linear in the amount of output desired[55]. Let the cost of producing y be $c(y) = cy$ where $c > 0$.

However, now consider the creation of the blueprint for this type of factory. Suppose it costs F to develop this blueprint. The cost of producing an amount y is now $\tilde{c}(y) = F + cy$. That is, if the firm spends less than F, it never manages to develop a factory, and cannot produce anything at all.

Suppose the firm spends a total amount X on production. For values of $X < F$, the firm fails to develop the blueprint, and produces nothing at all. For values of $X > F$, the firm successfully develops a new blueprint and spends an amount $X - F$ on producing goods. Given a cost c per unit, the number of goods produced is $y = \frac{\max\{X-F,0\}}{c}$.

Suppose $X > F$. What does doubling inputs to $2X$ do to output? Note that

$$y(2X) = \frac{2X - F}{c} = \frac{F}{c} + 2\left(\frac{X-F}{c}\right) = \frac{F}{c} + 2y(X) > 2y(X)$$

so doubling inputs *more than* doubles the output produced! Intuitively, the fixed cost of developing the blueprint has to be spent only once - once spent, the blueprint is nonrivalrous, and the firm can duplicate the production process embodied in it endlessly.

7.5.2 A Simple Romer Model

The insight that research is nonrival is at the heart of the Romer Model, but the way it shows up is subtle. We now study a version of the Neoclassical Growth Model that incorporates productivity growth through innovation.

The economy has two sectors, a production sector and a research sector. The production sector employs $N_{pt} = (1 - s_R)N_t$ workers and the research sector employs $N_{rt} = s_R N_t$ workers. Let A_t be a measure of the current state of ideas, then research adds to this measure of the stock of ideas according to the law of motion

$$A_{t+1} = A_t + Z_R N_{rt} A_t^{\phi}$$

where $\phi \leq 1$ is a parameter. We will explore the role of ϕ in detail below. Z_R is a parameter that governs the productivity of research over time.

Let's turn to the production sector, where the production function is assumed to be

$$Y_t = A_t K_t^{\alpha} N_{pt}^{1-\alpha}$$

where K_t, N_{pt} are aggregate inputs of capital and *production* labor. In specifying the production function this way, we identify TFP as equivalent to the measure of the stock

[55] To see this, suppose labor and capital cost w, r respectively. Then the cost of producing y is clearly $c(y) = wl + rk$. We argued above that doubling inputs leads to double the output. Thus, the cost of producing double the output, $2y$, is just $w \times 2l + r \times 2k = 2c(y)$. Thus $c(y)$ grows linearly in y

of ideas. Here is where non-rivalry shows up - irrespective of which firm is operating this technology, the state of knowledge is common to all firms.

We skip the details of the Romer Model[56] and move directly to characterizing a balanced growth path. Note that given s_R, output is

$$Y_t = A_t K_t^{\alpha} (1 - s_R)^{1-\alpha} N_t^{1-\alpha}$$

which implies that output per capita can be rewritten as

$$\frac{Y_t}{N_t} = A_t^{\frac{1}{1-\alpha}} \left(\frac{K_t}{N_t}\right)^{\alpha/(1-\alpha)} (1 - s_R) \implies g_{Y/N} = \frac{g_A}{1-\alpha}$$

Observe that the growth of A_t satisfies

$$g_{At} = \frac{A_{t+1} - A_t}{A_t} = Z_R s_R N_t A_t^{\phi - 1}$$

We now consider two cases.

- $\phi = 1$, the original case studied by Paul Romer. In this case, we have

  $$g_{At} = \frac{A_{t+1} - A_t}{A_t} = Z_R s_R N_t$$

  Observe that the *growth rate* of the economy is proportional to the *level* of the population of the economy. This *scale effect* is a consequence of the assumption that the share of researchers in the economy is a constant, but also shows up in more complex models where individuals endogenously choose whether to become researchers or not. This suggests that a higher population corresponds to a higher growth rate, which is counterfactual when applied across countries (i.e. it is not the case that more populous countries necessarily grow faster.). This is not necessarily evidence against the Romer model, since ideas flow across country borders. The appropriate way to think of the Romer model is to think of the entire world economy as an integrated source of ideas. Indeed, there is some evidence that over the long run of history, the rise in population levels has coincided with more rapid economic growth.

- $\phi < 1$, a case named the Jones Model after Charles Jones (who teaches at the GSB!). In this case,

  $$g_{At} = \frac{A_{t+1} - A_t}{A_t} = \frac{Z_R s_R N_t}{A_t^{1-\phi}}$$

  which implies that over time, as A_t grows, the growth rate of A falls as well. One interpretation of this case is that the productivity of each researcher falls as the set of ideas grows larger - in a sense, ideas get harder to find over time. Indeed, for values of $\phi > 0$ and $\phi < 1$, the past stock of ideas is useful in future research -

[56] See the hard questions document!

so researchers are "standing on the shoulders of past giants" - but the marginal contribution of the existing stock of ideas falls over time. When $\phi < 0$, a larger existing stock of ideas actually contributes negatively to research effort, a situation referred to as the "stepping on toes" case (think of the set of ideas becoming so large that research effort ends up being increasingly duplicated as the scope for original ideas falls.)

When $\phi < 1$, the growth rate of TFP will eventually stabilize at a level g_A, at which the growth in TFP each period exactly offsets the marginal decline in the productivity of research due to the increase in A_t. To find this point, start with the law of motion for TFP,

$$g_{At} = \frac{Z_R s_R N_t}{A_t^{1-\phi}}$$

Take logs on both sides, and then use the relationship between growth rates and logs to get

$$g_{g_{At},t} = g_N - (1-\phi) g_{At}$$

where $g_{g_{At},t}$ is the growth rate of the growth rate of A at date t and g_N is the growth rate of the population. In the long run, g_{At} is constant at g_A, so $g_{g_{At},t} = 0$. Thus, we have

$$g_A = \frac{g_N}{1-\phi}$$

7.6 Growth Accounting and Development Accounting

We have seen a number of different theories on long-run growth, and now it's time to evaluate these theories using the data. As a first pass, it is worth considering how to account for growth in the economy. One popular approach to doing this is to look within countries over time and try to parse out whether TFP growth, capital deepening, or human capital accumulation accounts for the majority of growth. This approach is called **growth accounting**.

Start with a simplified model that nonetheless nests most of the models we have studied. Suppose output is produced according to the production function

$$Y_t = A_t K_t^{\alpha} ((1 - s_E - s_R) h_t N_t)^{1-\alpha}$$

where A_t is TFP, K_t is capital, N_t is total population, s_E is the fraction of the population accumulating human capital through education, s_R is the fraction of the population involved in research activities, and h_t is human capital per worker. We assume that TFP growth g_A is constant, and given by

$$g_A = \frac{Z_R s_R N_t}{A_t^{1-\phi}} \implies A_t = \left(\frac{Z_R s_R N_t}{g_A} \right)^{1/(1-\phi)}$$

The Cobb-Douglas production function can be rewritten in the intensive form, as

$$\frac{Y_t}{N_t} = A_t^{1/(1-\alpha)} \left(\frac{K_t}{Y_t} \right)^{\alpha/(1-\alpha)} (1 - s_E - s_R) h_t$$

Substituting for A_t,

$$\frac{Y_t}{N_t} = \left(\frac{K_t}{Y_t}\right)^{\alpha/(1-\alpha)} (1 - s_E - s_R) h_t \left(\frac{Z_R s_R N_t}{g_A}\right)^{1/(1-\phi)}$$

Take logs and group terms, to get

$$\begin{aligned}
\log Y_t/N_t = const. + & \underbrace{\frac{\alpha}{1-\alpha} \log\left(\frac{K_t}{Y_t}\right)}_{\text{“Solow” term (Capital Deepening)}} \\
& + \frac{1}{1-\alpha}\left[\underbrace{\log h_t}_{\text{“Lucas” term (human capital accumulation)}}\right. \\
& + \underbrace{\frac{1}{1-\phi}\log s_R}_{\text{“Romer” term (rising share of researchers)}} \\
& \left. + \underbrace{\frac{1}{1-\phi}\log N_t}_{\text{“Jones” term (rising population)}}\right]
\end{aligned}$$

Empirically, the Solow term accounts for virtually nothing when it comes to post-war per-capita income growth in the US. The Lucas and Jones terms account for about 20% each, and the remaining 60% comes from the Romer term. This is manifested in a massive increase in research inputs in the US. Note that this mode of increasing output per worker is not sustainable, since s_R has an upper bound. Further, increases in s_R in the Jones model do not lead to permanent increases in the long-run growth rate of per capita income, only to level increases in output per worker.

Another approach to understanding what is important for growth is to look across countries at a given point in time, that is, to look at the *cross-section* of output per worker. Consider the simple Cobb-Douglas production function in intensive form,

$$\frac{Y_t}{N_t} = A_t^{1/(1-\alpha)} \left(\frac{K_t}{N_t}\right)^{\alpha/(1-\alpha)}$$

Consider two countries, the US and Sri Lanka. Suppose both countries operate the same technologies, but have different levels of TFP, labor and capital. This production function implies that

$$\frac{Y_{US,t}/N_{US,t}}{Y_{SL,t}/N_{SL,t}} = \left(\frac{A_{US,t}}{A_{SL,t}}\right)^{1/(1-\alpha)} \left(\frac{K_{US,t}/N_{US,t}}{K_{SL,t}/N_{SL,t}}\right)^{\alpha/(1-\alpha)}$$

Thus, differences in per-capita income between the US and Sri Lanka must come from differences in TFP or differences in the capital/labor ratios in the two countries.

Consistent with a Solow-like view of the world, globally, dispersion in capital/labor ratios is much smaller than dispersion in productivity. Dispersion in K/Y accounts for about 5% of differences in Y/N, while differences in A contribute the remaining 95%, of which human capital per worker accounts for about 45% and the remaining 50% are driven by technology differences and the misallocation of resources across firms. Human capital per worker can in turn depend on the quantity and quality of education acquired by workers in a country, the extent to which workers accumulate skills on the job, and worker health.

An important concern when studying the sources of growth globally is the issue of convergence vs divergence: whether cross-region differences in $\log Y/Pop$ are falling or rising over time. In the data, episodes of convergence include the experiences of the US states after the Civil War, the experiences of rich nations in the post-WWII world, and the rapid catch-up of poor countries since 2000 at least. However, the data also reveal episodes of divergence, such as between Asia and Europe between 1800-1950. Curiously, between 1960 and 2000, the data appear to show divergence - but once countries are weighted by their population, the data show convergence, driven by the rapid growth of emerging Asia, China and India.

7.7 The Future of Economic Growth

US TFP growth averaged around 2.9% a year between 1995 and 2005, but has since been disappointingly low. There are reasons to be pessimistic about US growth going forward, as the economy faces severe headwinds. These include

- Demographic challenges, including declines in population growth and declining fertility and cross-country mobility as well as in high-skill migration
- A leveling-off of educational attainment
- Declines in business dynamism since at least the 1980s, with declines in rates of firm entry and employment at young firms. This has been accompanied by falling rates of job reallocation (the sum of job creation and destruction). One possible explanation for this is that monopoly power and increasing returns to scale have led to the creation of large firms and high barriers to entry in many industries.
- Structural change in the US economy and accompanying "cost disease", as the US economy increasingly directs spending towards sectors with slow TFP growth.

However, there are reasons for optimism as well. As in the early 1990s, the world is seeing the rapid adoption of a new wave of technologies, including advanced robotics and AI systems and advances in biotechnology. In addition, huge portions of the world economy are still engaged in "catch-up" growth, and as these economies approach the global technology frontier, they will spur growth both through their demand for goods and services and through increasing the supply of high-skilled talent.

7.8 Growth and Misallocation

So far, we have studied models in which capital and labor are efficiently allocated across firms. An important consequence of this (which you will show in the problem set!) is that net marginal products of capital are equated across firms. Intuitively, suppose that net marginal products of capital were not equal across firms, and that there are two firms i, j such that $MPN_i > MPN_j$. Allocating a unit of labor away from firm i toward firm j reduces firm i's output by MPN_i and raises firm j's output by MPN_j. The net impact on total output is $MPN_i - MPN_j > 0$. But this means that we can raise total output in society relative to the original allocation, which means that the original allocation must not have been efficient to begin with.

The argument above implies that misallocation depresses aggregate productivity, and thus reduces TFP. It also implies that a good way to assess allocative efficiency is by considering the extent to which marginal products of capital and labor differ across firms. In the data, economies like China, India and Mexico do indeed seem to have much more widely dispersed marginal products of labor and capital.

What can cause an inefficient allocation of factors across firms? Some explanations proposed in the literature include

- Heterogeneous Market power, which causes some firms to have high marginal products of labor and capital level to the efficient level.
- Excessive regulation of firms, which can include regulations affecting firm entry and exit or regulations on labor mobility such as high firing costs. When entry barriers are high, low MPK or MPN firms survive in the marketplace more easily and more efficient entrants are unable to enter. High firing costs can also leave low MPN firms stuck with excess labor relative to what is efficient.
- The presence of state-owned enterprises, particularly state-owned banks, can lead to misallocation if they are on average less productive than private sector firms.
- Restrictions on Trade and international finance (such as barriers to FDI) tend to raise misallocation by allowing low productivity domestic firms to survive, hindering the growth of more efficient firms which could export.

8 The Real Business Cycle Model

So far, we have applied our neoclassical model to study long-run growth. We now turn to studying the business cycle, the phenomenon wherein macroeconomic aggregates appear to display persistent co-movements with each other. We will start by discussing some regularities in the cyclical behavior of macroeconomic variables. We will then discuss real business cycle models.

A real business cycle model is a model explaining business cycle co-movements as a result of fluctuations in TFP in a neoclassical growth model with perfectly competitive markets and fully flexible prices. We will see that, perhaps surprisingly, fluctuations in TFP in a neoclassical growth model produce co-movements in output, consumption and investment that are both qualitatively and quantitatively consistent with the data. We will then explore some difficulties faced by the RBC model in explaining co-movements in response to demand shocks, and explore extensions of the RBC model that make it appropriate for the study of monetary policy.

8.1 Business Cycle Facts

We first begin with a documentation of the facts governing the cycle. The stability of these regularities are what motivate the search for a theory of the business cycle.

First, consider output per worker. Over the long run, output per worker has grown at a roughly stable 2% a year, but this has recently slowed down. Between 1952 and 2000, output per person rose about 2.2% a year, but between 2000 and 2020, it rose at only about 1.2% a year. Output fluctuates around this trend.

The **cyclicality** of a macroeconomic aggregate is the extent to which it co-moves with output per capita over the business cycle. Macroeconomic variables can be

- **Procyclical**, meaning they positively co-move with output per capita (i.e. they rise during expansions and decline in recessions). Examples include consumption, investment, hours worked and the firm entry rate.
- **Acyclical**, meaning they display no statistically significant correlation with the cycle. These include the installed capital stock, government spending, the real interest rate and the real wage rate.
- **Countercyclical**, meaning they negatively co-move with output per capita (i.e. they fall in expansions and rise in recessions). These include the unemployment rate and the firm exit rate.

While the cyclicality of macroeconomic aggregates is important, business cycle modeling is also influenced by the relative size of the correlations between output per capita and macroeconomic variables. These "second moments" are useful to identify mechanisms and discriminate between models we will use to study the cycle. Some first-order facts documented in the data are the following.

- **Consumption and Investment**: In the data, consumption is less cyclical than output, but investment is considerably more cyclical than output is: A 1% increase in output growth is associated with a 0.55% increase in consumption growth but a 3.75% increase in investment growth! At the same time, total capital is acyclical. This is because the size of total capital stock is much larger than the size of output or investment at any given date.
- **Hours worked:** While employment rates are highly cyclical, hours worked per employed worker are much less cyclical, indicating that movements in total labor input (i.e. total hours worked) is driven much more by the *extensive* margin of employment than the *intensive* margin of how many hours each employed worker works.
- **Unemployment Rates:** Changes in national unemployment rates are strongly negatively related to output growth: a 1% higher unemployment rate is associated with a 1.6% lower output growth rate, a fact known as Okun's Law. National unemployment rates and state and local unemployment rates have a strong tendency to co-move.
- **Government Spending:** Government spending tends to co-move only weakly with output.

8.2 The Baseline Real Business Cycle Model

8.2.1 An Old View of Business Cycles

Prior to modern business cycle theory, there were a number of attempts to rationalize the business cycle as the outcome of completely deterministic dynamic models of the economy. This line of research argued that the data could be viewed as a composition of several long-run cycles of roughly fixed periodicity, such as the "Kondratiev cycle" with lengths of 45-60 years emphasizing technological changes and the "Kitchin cycles" in inventory with much shorter periods of 3-5 years. It attempted to identify these cycles using time series analysis and then associate them with economic forces. These models typically featured second order difference equations, and argued that particular solutions to these equations which featured cyclical behavior were consistent with the data.

The modern view of business cycles argues that there is limited evidence for the idea that the data should be thought of as a composition of many long-term cycles. Instead, the modern view of business cycles emphasizes the importance of shocks to the economy and the mechanisms through which these shocks are amplified and propagated through an economy. In a modern business cycle model, certain variables are taken as exogenous, and changes in exogenous variables are treated as shocks. The content of a model is a description of how these shocks then affect all other macroeconomic aggregates through the economic relationships in the model.

8.2.2 The Model's Goal

Before describing the RBC model, it's worth remembering what we're aiming for here. We are NOT looking for a complete theory of the business cycle that explains every co-movement and every aspect of the business cycle. Rather, we're looking for the *simplest* model that can achieve the following goals.

- The model must be a consistent description of an economy in which agents understand the nature of uncertainty they face and take rational decisions based on available information, making use of rational expectations. That is, the model must
  - take explicit account of the **D**ynamic and **S**tochastic nature of the macroeconomy
  - be a **G**eneral **E**quilibrium model, in which prices and quantities adjust so that all markets are in equilibrium.
- The model's **quantitative** implications for macroeconomic aggregates must be consistent with the data. The target for our model is going to be its ability to match the following facts.
  - Consumption and Investment are procyclical, with consumption volatility about 60% that of output and investment volatility about 3 times that of output.
  - The Solow Residual and total hours worked are procyclical.
  - Recessions are persistent and recoveries from them can be slow.

8.2.3 The Model in Words

Put quite simply, the RBC model postulates that if consumers have a concave felicity function and are forward looking, then persistent fluctuations in TFP, often identified with shocks to technology, can generate movements in macroeconomic aggregates that are consistent with the facts above. Here's how the model tackles the facts above.

- The procyclicality in the Solow Residual is the key mechanism generating business cycles, and the model will not explain the source of these fluctuations. Another way to put this: TFP is exogenous in the RBC Model.
- With concave preferences, consumers will engage in consumption smoothing, which dampens fluctuations in consumption relative to those in output and investment.
- With quantitatively realistic fluctuations in TFP and a model calibrated to be consistent with the US capital-output ratio, fluctuations in the marginal product of capital and hence in the target capital-output ratio are small, but still induce large changes in investment.

8.2.4 The Model

The real business cycle model is essentially a neoclassical growth model, with three important changes:

- Hours worked will now be an endogenous choice made by households, like in the two-period model we studied in section 5.2.
- We will abstract from economic growth: that is, productivity will not grow at a constant rate, and there will be no population growth. This does not change the model's implications for aggregates, but requires us to be careful when we try to match the model to the data.
- TFP is going to be **stochastic**: that is, each period, the exact level of TFP that will prevail in the next period is not certain.

As in the neoclassical growth model, households solve the intertemporal problem below:

$$\max_{\{C_t,N_t\}_{t=0}^{\infty}} \mathbb{E}_0 \sum_{t=0}^{\infty} \frac{1}{(1+\rho)^t}\left[\frac{C_t^{1-1/\sigma}}{1-1/\sigma} - \frac{N_t^{1+1/\psi}}{1+1/\psi}\right]$$

subject to

$C_t + K_{t+1} = (1+r_t)K_t + w_t N_t \ \ \forall\, t$

K_0 given

A condition to rule out explosive solutions.

As before, the solution to this problem is characterized by the intratemporal Euler condition

$$N_t^{1/\psi} = C_t^{-1/\sigma} w_t \tag{20}$$

the intertemporal Euler equation

$$C_t^{-1/\sigma} = \frac{1+r_t}{1+\rho} C_{t+1}^{-1/\sigma} \tag{21}$$

and the sequence of budget constraints and the condition to rule out explosive solutions.

Firms solve a problem completely analogous to the one they solved in the Neoclassical growth model - they rent capital at rental rate r_t from households and hire labor at a wage rate of w_t. They produce output according to the Cobb-Douglas technology

$$Y_t = A_t K_t^{\alpha} N_t^{1-\alpha}$$

They solve the static problem

$$\max_{K_t,N_t} A_t K_t^{\alpha} N_t^{1-\alpha} - \delta K_t - w_t N_t - r_t K_t$$

where δ is the depreciation rate and A_t is the level of TFP. The solution to this problem is characterized by the two first-order conditions

$$MPK_t \equiv \alpha \frac{Y_t}{K_t} = \alpha \left(\frac{K_t}{Z_t N_t} \right)^{\alpha - 1} = r_t + \delta \tag{22}$$

$$MPN_t \equiv (1-\alpha) \frac{Y_t}{N_t} = (1-\alpha) Z_t \left(\frac{K_t}{Z_t N_t} \right)^{\alpha} = w_t \tag{23}$$

Note that the firm takes TFP and prices as given when it makes its choices, and that its problem is effectively static. We will assume that the log of TFP follows an Autoregressive Process: that is, we have

$$\log A_{t+1} = \rho \log A_t + \varepsilon_{t+1} \tag{24}$$

where ε_t is drawn each period from a distribution with mean 0 and variance σ^2. Most quantitative work assumes that ε is normally distributed.

Finally, the model is closed with market clearing conditions. First, labor markets must clear, so labor demand and labor supply must be equal. Second, capital demand must equal the pre-installed capital stock available at date t. Finally, total output must either be consumed or invested in capital, so that $Y_t = C_t + K_{t+1} - (1-\delta)K_t$.

Given initial values A_0, K_0, an **equilibrium** of the RBC Model is a collection of processes for macroeconomic aggregates $\{A_t, Y_t, K_t, N_t, C_t\}$ and prices $\{w_t, r_t\}$ satisfying the following conditions.

- A_t satisfies the law of motion for A_t, equation 24.
- C_t, N_t satisfy the household's Euler equations 20 and 21, and its budget constraint at all dates, given the prices.
- K_t, N_t satisfy the firm's optimality conditions, given the prices.
- All markets clear.

In an equilibrium of the RBC Model, the paths of the aggregate variables therefore satisfy the set of dynamic equations

$$\begin{aligned}
\log A_{t+1} &= \rho \log A_t + \varepsilon_{t+1} \\
N_t^{1/\psi} &= C_t^{-1/\sigma} w_t \\
C_t^{-1/\sigma} &= \frac{1+r_t}{1+\rho} C_{t+1}^{-1/\sigma} \\
w_t &= A_t (1-\alpha) K_t^{\alpha} N_t^{-\alpha} \\
r_t + \delta &= A_t \alpha K_t^{\alpha-1} N_t^{1-\alpha} \\
Y_t &= A_t K_t^{\alpha} N_t^{1-\alpha} \\
Y_t &= C_t + K_{t+1} - (1-\delta) K_t
\end{aligned}$$

given values for A_0, K_0.

Note that the equilibrium is *dynamic* - the paths Y_t, C_t, K_t, N_t will not in general be constant over time. This is because the shocks ε are constantly hitting the economy, and inducing changes in behavior by the household and the firm in response to productivity changing. However, note that in equilibrium, if we turn off productivity shocks, all variables will indeed just be constant over time. Second, note that the dynamics of aggregates in the model are driven by dynamics in two variables: TFP and capital. When $\rho > 0$, a one-period shock to TFP (i.e. a higher ε for one period) leads to a higher level of TFP for a prolonged period of time. Similarly, as long as $\delta < 1$, a one-period jump in investment raises the capital stock for a prolonged period of time as well.

8.3 The Steady State of the RBC Model

How can we use the RBC model to study business cycles? The approach we follow is as follows. We start by considering the steady state of the RBC Model, defined as a model equilibrium in which A_t is constant at its long run mean value (so the ε_t are all 0). We then study the behavior of the model in response to a shock to ε, sometimes called the *innovation* in A_t.

First consider the steady state. Start with the law of motion for A_t. If $\varepsilon_t = 0$ for all dates t then clearly there is nothing in the economy that changes over time, so we must have $A_t = A$ for all dates t. From the law of motion for A, we must have $\log A = \rho \log A \implies \log A = 0$, which implies that $A = 1$ at all dates.

In a steady state, we also know that the aggregates Y, C, K, N and the prices w, r are constant. Given that $A = 1$ at all dates, the system of equations above becomes

$$
\begin{aligned}
N^{1/\psi} &= wC^{-1/\sigma} \\
C^{-1/\sigma} &= \frac{1+r}{1+\rho} C^{-1/\sigma} \\
w &= (1-\alpha) K^{\alpha} N^{-\alpha} \\
r + \delta &= \alpha K^{\alpha-1} N^{1-\alpha} \\
Y &= K^{\alpha} N^{1-\alpha} \\
Y &= C + K - (1-\delta) K
\end{aligned}
$$

which is a system of 6 equations in 6 endogenous variables. We can solve this system systematically as follows.

- Start with the intertemporal Euler Equation. Canceling the $C^{-1/\sigma}$ from both sides immediately gives $r = \rho$.
- The firm's FOC for capital then gives us the equilibrium K/N ratio,

$$
\frac{K}{N} = \left(\frac{\rho + \delta}{\alpha} \right)^{1/(\alpha-1)}
$$

Note that this ratio is a function of exogenous variables and parameters only. Given this, we know that the wage rate, from the firm's FOC for labor, is

$$w = (1-\alpha)\left(\frac{\rho+\delta}{\alpha}\right)^{\alpha/(\alpha-1)}$$

- Consider the goods market clearing condition and the production function. We have,

$$\begin{aligned}
Y &= C + \delta K \\
\implies \frac{K^{\alpha}N^{1-\alpha}}{N} &= \frac{C}{N} + \delta\frac{K}{N} \\
\implies \frac{C}{N} &= \left(\frac{K}{N}\right)^{\alpha} - \delta\frac{K}{N} \\
&= \left(\frac{\rho+\delta}{\alpha}\right)^{\alpha/(\alpha-1)} - \delta\left(\frac{\rho+\delta}{\alpha}\right)^{1/(\alpha-1)}
\end{aligned}$$

- Consider the labor supply FOC. We have,

$$\begin{aligned}
N^{1/\psi} &= wC^{-1/\sigma} \\
\implies N^{1/\psi+1/\sigma} &= w\left(\frac{C}{N}\right)^{-1/\sigma} \\
\implies N^{1/\psi+1/\sigma} &= (1-\alpha)\left(\frac{\rho+\delta}{\alpha}\right)^{\alpha/(\alpha-1)} \times \left[\left(\frac{\rho+\delta}{\alpha}\right)^{\alpha/(\alpha-1)} - \delta\left(\frac{\rho+\delta}{\alpha}\right)^{1/(\alpha-1)}\right]^{-1/\sigma}
\end{aligned}$$

Raising both sides to the power $1/(1/\psi+1/\sigma)$, we get

$$N = \left\{(1-\alpha)\left(\frac{\rho+\delta}{\alpha}\right)^{\frac{\alpha}{1-\alpha}} \times \left[\left(\frac{\rho+\delta}{\alpha}\right)^{\frac{\alpha}{\alpha-1}} - \delta\left(\frac{\rho+\delta}{\alpha}\right)^{\frac{1}{\alpha-1}}\right]^{-1/\sigma}\right\}^{\frac{1}{\frac{1}{\psi}+\frac{1}{\sigma}}}$$

While this expression is forbidding, note that the right hand side is a function of exogenous variables and parameters only. We have solved for employment in the *general equilibrium* of the RBC model - in closed form!

- Given N and the formulas for C/N and K/N, we can calculate C and K easily. Output can then be calculated using the production function.

8.4 Dynamics in the RBC Model

To study the dynamics of variables in the RBC model, a common approach is to study *Impulse responses*, which trace out the response of variables to a TFP shock. While the calculation of impulse responses is well beyond the scope of Econ 52, we will explore the qualitative shapes associated with them.

8.4.1 A Transitory Productivity shock

Suppose the economy is at a steady state. At date $t = 0$, TFP rises for one period and then returns to its steady-state level at $t = 1$ and stays there forever. This change is not anticipated. What happens to the remaining variables?

- Consider k_t. Since capital is pre-determined in the RBC model - so k_0 was installed at $t = -1$ - it does not respond to the TFP shock at all.
- If hours worked didn't respond to the shock, the result would be an increase in output at date $t = 0$, which should induce an increase in consumption. Since agents want to smooth consumption over time, savings increase at date $t = 0$, which raises investment at date $t = 0$ as well. The increase in investment means that capital at dates $t = 1, 2, \ldots$ will be higher for a while, as the increase due to investment at date $t = 1$ slowly depreciates. Taken together, this means that output will also be higher for a while.
- The productivity shock has offsetting income and substitution effects on labor supply at date $t = 0$: on the one hand, the higher TFP level means higher labor productivity, but on the other, the higher income level for a while may raise $PVLR$ by enough to create an offsetting income effect. The typical calibration of an RBC model leads substitution effects to dominate in the short run, which means that hours worked typically increase at date $t = 0$. The overall rise in hours is, however, dampened by income effects.
- At dates $t = 1, 2, \ldots$, productivity is back to its original level, so only the income effect survives. Thus, hours worked are lower at all dates $t = 1, 2, \ldots$. Since capital is higher but hours worked are lower, output is theoretically ambiguous at all dates $t = 1, 2, \ldots$. Note that $PVLR$ must rise overall nonetheless, for the income effect to be in the right direction.

The typically calibrated RBC model thus predicts that in response to a positive TFP shock, consumption, investment and hours worked all increase, as in the data. Due to consumption smoothing, it also predicts that consumption increases by less than output. Note that in the case of a one-period increase in productivity, the increase in investment is driven entirely by the increased desire of households to save, and not at all by a desire by firms to take advantage of the higher productivity level - firms

8.4.2 A Persistent Change in Productivity

Now suppose that the economy, starting in the steady state, experiences an increase in TFP at date $t = 0$ which decays only slowly. Again, this change is not anticipated at the start of date $t = 0$.

- The analysis of the behavior of consumption and hours for $t = 0$ above is still valid. Note that the income effect on hours worked will now be much stronger at date $t = 0$, since productivity will be higher for a long time and so will output, all else equal. The typical RBC calibration nonetheless tends to ensure that hours worked increase on impact, i.e. that the substitution effect dominates. By contrast to the previous case, investment at date $t = 0$ increases for two reasons - first, for consumption smoothing, and second, in order to take advantage of the higher productivity at date $t = 1$.

- Consider dates $t = 1, 2, \ldots$. Since productivity is higher at all of these dates, investment will also be higher than its steady state level. As productivity falls back to its steady state level, investment will also decline. At all of these dates, hours worked are technically ambiguous - there are offsetting income and substitution effects.

The typically calibrated TFP process features a very high persistence of shocks, required both to obtain the strong substitution effects necessary for an increase in labor supply and in order to match the relatively long durations of business cycles.

8.5 Shortcomings of the RBC Framework

The RBC model faces some key challenges.

- The model's internal propagation mechanisms are weak. There are only two dynamic equations in the model, the transition law for TFP and the law of motion for capital. Over the business cycle, movements in capital are small. This is related to the fact that the capital stock is large relative to output, so in order to move K by a lot would require enormous swings in investment.
- The model's theory of employment can seem strange. In an RBC model, the fall in hours worked in a recession is driven by substitution effects: in periods when TFP is low, the lower returns to working induce agents to consume leisure rather than supply labor. There is no notion of "involuntary" unemployment in an RBC model.
- In typical calibrations of the RBC model, fluctuations in hours worked are relatively small while fluctuations in wages are large, while the opposite is true in the data. This is driven by the fact that offsetting income and substitution effects make labor supply relatively inelastic with respect to the real wage.
- In the data, expansions of government spending tend to lead to expansions in economic activity on impact with increases in both consumption and labor. In a neoclassical model, however, this is not typically what happens, and the logic for why can be seen from the static first order condition for labor. An increase in the demand for output raises labor demand on impact, and for labor markets to clear, agents must be willing to work harder. However, from the labor supply first order condition, households will only be willing to work harder - implying a higher marginal disutility of labor - only if their marginal utility of consumption is higher. But this requires consumption to *fall*. RBC models with frictional labor markets can solve some of these challenges.

9 Sticky Prices and Business Cycles

By construction, the RBC model features complete short-run flexibility in prices and quantities. Under the standard calibration of these models, prices are particularly volatile and quantities not as much. This is particularly true of the labor market - due to offsetting income and substitution effects in response to shocks to TFP, hours worked do not vary much quantitatively in simulations of the model, while wage rates required to clear the labor market do.

In the data, prices and wages do not appear to change regularly - in the US, the typical consumer price changes every 8 months or so if we exclude temporary price discounts. Wages change, on average, just about once a year. Further, employment is particularly volatile over the business cycle, and baseline RBC models emphasizing flexible labor markets have a hard time matching this.

9.1 Money and Prices

9.2 Why Sticky Prices allow Monetary Policy to Work

9.3 Monetary Policy: Details

References

Lucas, Robert and Thomas Sargent, "After Keynesian macroeconomics," *Quarterly Review*, 1979, *3* (Spr).

Lucas, Robert E, "Understanding business cycles," *Carnegie-Rochester Conference Series on Public Policy*, 1977, *5*, 7–29.

Parker, Jeffrey, "Macroeconomics: Modeling the Behavior of Aggregate Variables," in "Economics 314 Coursebook, 2010" 2010, pp. 1–32.

Ramsey, F. P., "A Mathematical Theory of Saving," *The Economic Journal*, 1928, *38* (152), 543–559.

Romer, Paul M., "Endogenous Technological Change," *Journal of Political Economy*, 1990, *98* (5), S71–S102.

Stevenson, Betsey and Justin Wolfers, "Economic Growth and Subjective Well-Being: Reassessing the Easterlin Paradox," *Brookings Papers on Economic Activity*, 2008, *1* (Spr).